



Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas/Computación.

*Facultad de la Energía, las Industrias y los Recursos Naturales No Renovables*

---

## CARRERA DE INGENIERÍA EN SISTEMAS

# Minería de Datos Educativa aplicada al descubrimiento de patrones en registros académicos de estudiantes de Ingeniería con asignaturas relacionadas a las matemáticas

TESIS DE GRADO PREVIA A LA  
OBTENCIÓN DEL TÍTULO DE  
INGENIERO EN SISTEMAS

### Autor:

- Jhulissa Isabel Villamagua Poma

### Director

- Ing. Luis Antonio Chamba Eras, PhD.

LOJA – ECUADOR

2021

# CERTIFICACIÓN

Certificado Nro. 008-CIS-FEIRNNR-UNL

Luis Antonio Chamba Eras, **DIRECTOR DEL TRABAJO DE TITULACIÓN.**

## CERTIFICA:

Que la egresada: **JHULISSA ISABEL VILLAMAGUA POMA**, con C.I: **1150692414**, aprobó la asignatura de Trabajo de Titulación en el 10 Ciclo correspondiente al período académico: 04-05-2020 al 28-08-2020; respecto del desarrollo de su Trabajo de Titulación denominado **“MINERÍA DE DATOS EDUCATIVA APLICADA AL DESCUBRIMIENTO DE PATRONES EN REGISTROS ACADÉMICOS DE ESTUDIANTES DE INGENIERÍA CON ASIGNATURAS RELACIONADAS A LAS MATEMÁTICAS”**, opción de titulación, escogida dentro del periodo académico de culminación de sus estudios, siendo las **13H00** del **20** de **agosto** del **2021**, se certifica que se ha cumplido con el cien por ciento (100%) del mismo y está en condiciones de continuar con los procesos administrativos que correspondan.

Loja, **20 de agosto** del **2021**



Firmado electrónicamente por:  
**LUIS ANTONIO  
CHAMBA ERAS**

Ing. Luis Antonio Chamba Eras, Ph.D.  
**DIRECTOR DEL TRABAJO DE TITULACIÓN**

## **AUTORÍA**

Yo, **Jhulissa Isabel Villamagua Poma**, declaro ser el autor del presente Trabajo de Titulación y eximo expresamente a la Universidad Nacional de Loja y a sus representantes jurídicos de posibles reclamos o acciones legales, por el contenido de la misma

Adicionalmente acepto y autorizo a la Universidad Nacional de Loja, la publicación de mi trabajo de titulación en el Repositorio Institucional, Biblioteca Virtual.

**Firma:**



**Cédula:** 1150692414

**Fecha:** Loja, 25 de noviembre del 2021

## **CARTA DE AUTORIZACIÓN POR PARTE DEL AUTOR, PARA LA CONSULTA, REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO COMPLETO**

Yo, JHULISSA ISABEL VILLAMAGUA POMA declaro ser autor del trabajo de titulación que versa: “MINERÍA DE DATOS EDUCATIVA APLICADA AL DESCUBRIMIENTO DE PATRONES EN REGISTROS ACADÉMICOS DE ESTUDIANTES DE INGENIERÍA CON ASIGNATURAS RELACIONADAS A LAS MATEMÁTICAS”, como requisito para optar al grado de: INGENIERO EN SISTEMAS; autorizo al Sistema Bibliotecario de la Universidad Nacional de Loja para que, con fines académicos, muestre al mundo la producción intelectual de la Universidad, a través de la visibilidad de su contenido de la siguiente manera en el Repositorio Digital Institucional: Los usuarios pueden consultar el contenido de este trabajo en el (RDI), en las redes de información del país y del exterior, con los cuales tenga convenio la Universidad.

La Universidad Nacional de Loja, no se responsabiliza por el plagio o copia de la tesis que realice un tercero.

Para constancia de esta autorización, en la ciudad de Loja, a los 25 días del mes de noviembre del 2021.

**Firma:**



**Autor:** Jhulissa Isabel Villamagua Poma

**Cédula:** 1150692414

**Dirección:** Loja (Calle Éxodo de Yangana, Nro. 381-315)

**Correo Electrónico:** 3jhuvi@gmail.com

**Celular:** 0989106042

### **DATOS COMPLEMENTARIOS**

**Director de tesis:** Ing. Luis Antonio Chamba Eras, Ph.D.

**Tribunal de grado:** Ing. Óscar Miguel Cumbicus Pineda, Mg. Sc.

Ing. Andrés Roberto Navas Castellanos, Mg. Sc.

Ing. Franco Hernán Salcedo López, Mg. Sc.



## **DEDICATORIA**

A mis padres Agustín Villamagua y Rosario Poma, este triunfo es gracias a ustedes, pese a las circunstancias que se presentaron nunca se dieron por vencidos, ya que siempre estuvieron apoyándome para poder cumplir este sueño, día a día buscando la manera de poder sacarnos adelante a mis hermanos y a mí, gracias por todos sus esfuerzos y sacrificios papitos.

**Jhulissa Villamagua.**

## **AGRADECIMIENTO**

Infinitas gracias a Dios, por brindarme salud y la sabiduría necesaria para poder cumplir con éxito una meta más.

A toda mi familia, en especial a mis padres, por su infinito amor y su apoyo durante mi todos los aspectos, especialmente en mi formación académica.

A cada uno de mis hermanos, por todos sus consejos y apoyo para poder salir adelante.

A mi complemento Javier, quien ha sido mi apoyo en todo momento, gracias a todos sus consejos que me ayudaron a no darme por vencida y seguir luchando por mi sueño.

A mis amigos, quienes de una u otra manera aportaron a la culminación de este trabajo, gracias por sus consejos y apoyo incondicional.

A mi director del presente trabajo de titulación, el Ing. Luis Antonio Chamba Eras, quien con su experiencia y conocimiento me ha orientado de la mejor manera para poder culminarlo con éxito.

Finalmente, mi gratitud infinita a cada uno de los docentes que formaron parte de mi formación académica, gracias por brindarme su conocimiento y experiencia, para que sueños como este se hicieran realidad.

# ÍNDICE DE CONTENIDOS

CERTIFICACIÓN .....	II
AUTORÍA .....	III
CARTA DE AUTORIZACIÓN .....	IV
DEDICATORIA.....	V
AGRADECIMIENTO.....	VI
ÍNDICE DE CONTENIDOS .....	VII
ÍNDICE DE FIGURAS .....	XI
<b>1. TITULO.....</b>	<b>1</b>
<b>2. RESUMEN.....</b>	<b>2</b>
ABSTRACT .....	4
<b>3. INTRODUCCION.....</b>	<b>5</b>
<b>4. REVISIÓN DE LITERATURA.....</b>	<b>8</b>
4.1.    Deserción Universitaria .....	8
4.1.1.    Definición .....	9
4.1.2.    Factores de la deserción Universitaria .....	9
4.2.    Las Matemáticas.....	11
4.2.1.    Influencia de las Matemáticas en los estudiantes universitarios .....	11
4.3.    Inteligencia Artificial .....	12
4.3.1.    Definición de la Inteligencia Artificial .....	12
4.3.2.    Inteligencia Artificial en la Educación .....	12
4.4.    Minería de Datos.....	13
4.5.    Técnicas de Minería de Datos .....	15
4.5.1.    Técnicas Supervisadas o Predictivas .....	16
4.5.2.    Técnicas No Supervisadas o Descriptivas .....	16
4.6.    Herramientas Minería de Datos .....	16
4.6.1.    Weka.....	17

4.6.2.	RapidMiner.....	17
4.6.3.	Spyder - Python .....	17
4.6.4.	Pandas.....	18
4.7.	Metodología de Minería de Datos .....	18
4.7.1.	CRISP-DM .....	19
<b>5.</b>	<b>MATERIALES Y MÉTODOS.....</b>	<b>22</b>
5.1.	Contexto.....	23
5.2.	Proceso.....	23
5.3.	Recursos.....	24
5.3.1.	Recursos Científicos .....	24
5.3.2.	Recursos Técnicos .....	25
5.3.3.	Recursos Éticos .....	25
5.4.	Participantes .....	26
<b>6.</b>	<b>RESULTADOS .....</b>	<b>27</b>
6.1.	Fase 1: Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería. ....	27
6.1.1.	Establecer métricas y lineamientos para la selección de trabajos relacionados con el tema de estudio.....	27
6.1.2.	Obtener y analizar los trabajos relacionados en base a métricas y lineamientos antes mencionados.....	28
6.1.3.	Elaborar Documentación.....	31
6.2.	Fase 2: Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa. ....	32
6.2.1.	Seleccionar la herramienta de minería de datos a utilizar para el desarrollo del trabajo.....	32
6.2.2.	Determinar el algoritmo más óptimo para la minería de datos.....	33
6.2.3.	Analizar e interpretar los resultados obtenidos. ....	33

6.3.	Fase 3: Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación. ....	45
6.3.1.	Definir los objetivos del plan de acción. ....	46
6.3.2.	Construir el plan de acción de acuerdo a los objetivos planteados..	47
<b>7.</b>	<b>DISCUSIÓN</b> .....	<b>58</b>
7.1.	Desarrollo de la propuesta alternativa .....	58
7.1.1.	Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería .....	58
7.1.2.	Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa. ....	59
7.1.3.	Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación. ....	61
7.2.	Valoración técnica, económica y ambiental .....	62
7.2.1.	Valoración Técnica .....	62
7.2.2.	Valoración Económica .....	62
7.2.3.	Valoración Económica .....	64
7.3.	Limitaciones.....	64
7.3.1.	Utilizar nuevas técnicas de Minería de Datos .....	64
7.3.2.	Aplicabilidad de Minería de Datos a todas las asignaturas.....	64
7.3.3.	Población no contemplada.....	65
<b>8.</b>	<b>CONCLUSIONES</b> .....	<b>66</b>
<b>9.</b>	<b>RECOMENDACIONES</b> .....	<b>68</b>
	TRABAJOS FUTUROS .....	69
<b>10.</b>	<b>BIBLIOGRAFIA</b> .....	<b>70</b>
<b>11.</b>	<b>ANEXOS</b> .....	<b>83</b>
	Anexo 1. Revisión sistemática de la literatura.....	83

Anexo 2. Análisis Exploratorio de los datos .....	115
Anexo 3. Reporte de Minería .....	127
Anexo 4. Ingeniería en Sistemas – Primer ciclo.....	168
Anexo 5. Ingeniería en Sistemas – Segundo ciclo .....	181
Anexo 6. Ingeniería en Sistemas – Tercer ciclo.....	194
Anexo 7. Ingeniería en Sistemas – Cuarto ciclo .....	208
Anexo 8. Ingeniería en Sistemas – Quinto ciclo.....	222
Anexo 9. Ingeniería en Sistemas – Sexto ciclo .....	236
Anexo 10. Ingeniería en Sistemas – Séptimo ciclo .....	250
Anexo 11. Ingeniería en Sistemas – Octavo ciclo.....	263
Anexo 12. Ingeniería en Sistemas – Noveno Ciclo .....	278
Anexo 13. Ingeniería en Sistemas – Décimo ciclo .....	292
Anexo 14. Ingeniería en Computación – Primer ciclo .....	306
Anexo 15. Ingeniería en Computación – Segundo ciclo.....	322
Anexo 16. Ingeniería en Computación – Tercer Ciclos .....	337
Anexo 17. Ingeniería en Computación – Cuarto ciclo .....	352
Anexo 18. Certificación: Máster Universitario en Ingeniería Matemática y Computación.....	366
Anexo 19. Certificación: Miembro del Consejo Consultivo - UNL .....	367
Anexo 20. Certificación: Magíster en Docencia Universitaria e Investigación Educativa .....	368
Anexo 21. Anteproyecto de Titulación.....	369

## ÍNDICE DE FIGURAS

Figura 1. Factores de Deserción Universitaria .....	10
Figura 2. El ciclo de aplicación de minería de datos en sistemas educativos ..	14
Figura 3. Técnicas de Minería de Datos.....	16
Figura 4. Herramientas para Minería de Datos obtenidas de KDnuggets .....	17
Figura 5. Resultados de CRISP-DM en encuesta realizada por KDnuggets (2014) .....	18
Figura 6. Fases de la Metodología CRISP-DM .....	19
Figura 7. Desglose de las fases de la Metodología CRISP-DM .....	20
Figura 8. Diagrama de flujo para la obtención de documentos relacionados ...	30
Figura 9. Herramientas a utilizarse para la Minería de Datos Educativa. ....	33
Figura 10. Selección de variables para la elaboración del modelo.....	35
Figura 11. Reglas obtenidas aplicando filtros (Conf – Lift) .....	35
Figura 12. Reglas de asociación en RapidMiner .....	36
Figura 13. Reglas de asociación obtenidas por Python .....	37
Figura 14. Clústeres con sus respectivas instancias.....	37
Figura 15. Clústeres obtenidos por Weka. ....	38
Figura 16. Clúster con total de instancias .....	39
Figura 17. Agrupación de cada Instancia a su respectivo Clúster.....	40
Figura 18. Representación gráfica de los clústeres obtenidos .....	40
Figura 19. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	41
Figura 20. Relación Género – Tipo de Colegio .....	42
Figura 21. Relación Edad – Programación Avanzada.....	43
Figura 22. Relación Género - Programación Avanzada .....	43
Figura 23. Relación Estadística Inferencial - Programación Avanzada .....	44
Figura 24. Confianza - Quinto ciclo Sistemas.....	45
Figura 25. Búsqueda avanzada en la base de datos IEEE. ....	87
Figura 26. Técnicas de minería de datos .....	113
Figura 27. Herramientas para la ejecución de MD. ....	113
Figura 28. Metodologías para el proceso de desarrollo de Minería de Datos. ....	114
Figura 29. Variables totales dentro del conjunto de datos inicial.....	115
Figura 30. Carreras presentes en el conjunto de datos.....	118
Figura 31. Periodo lectivo dentro del conjunto de datos.....	119

Figura 32. Género de los estudiantes .....	120
Figura 33. Estado civil de los estudiantes .....	120
Figura 34. Etnia con la que se identifican los estudiantes.....	121
Figura 35. Sector de procedencia de los estudiantes.....	121
Figura 36. Nacionalidad a la que pertenecen cada estudiante.....	121
Figura 37. Provincia de nacimiento de los estudiantes .....	122
Figura 38. Estado de trabajo de los estudiantes. ....	122
Figura 39. Ingreso mensual de los estudiantes.....	123
Figura 40. Número de hijos bajo la responsabilidad de los estudiantes .....	123
Figura 41. Tipo de colegio en la que se formaron los estudiantes .....	124
Figura 42. Ciclos por los que los estudiantes deben pasar para culminar su Carrera. ....	124
Figura 43. Mallas Curriculares para los estudiantes.....	125
Figura 44. Estado de la matrícula perteneciente a los estudiantes. ....	126
Figura 45. Datos originales de los registros académicos .....	131
Figura 46. Estructura del conjunto de datos.....	140
Figura 47. Transposición de asignaturas únicas a filas.....	143
Figura 48. Interfaz principal de Weka.....	145
Figura 49. Interfaz de Weka para subir archivos a procesar. ....	146
Figura 50. Interfaz de RapidMiner .....	146
Figura 51. Crear reglas de asociación.....	147
Figura 52. Cargar dataset en RapidMiner .....	148
Figura 53. Conjunto de datos cargados .....	148
Figura 54. Implementar componente para transformar datos .....	149
Figura 55. Añadir componente FP-Growth.....	149
Figura 56. Resultados obtenidos por componente FP-Growth.....	150
Figura 57. Agregar Componente para Reglas de Asociación.....	150
Figura 58. Ejecutar algoritmo Simple KMeans .....	150
Figura 59. Configuración del Algoritmo KMeans .....	151
Figura 60. Algoritmo k-means en RapidMiner .....	151
Figura 61. Añadir Componente de Normalización.....	152
Figura 62. Añadir Componentes Normalización y KMeans .....	153
Figura 63. Agregar Componente Performance.....	153
Figura 64. Diseño completo para Clústering .....	154



Figura 65. Componente adicional Parse Numbers.....	154
Figura 66. Resultados de Prueba y Error .....	155
Figura 67. Confianza - Primer ciclo Sistemas.....	156
Figura 68. Confianza - Segundo ciclo Sistemas.....	157
Figura 69. Confianza - Tercer ciclo Sistemas.....	158
Figura 70. Confianza - Cuarto ciclo Sistemas .....	159
Figura 71. Confianza - Quinto ciclo Sistemas.....	160
Figura 72. Confianza - Sexto ciclo Sistemas.....	160
Figura 73. Confianza - Séptimo ciclo Sistemas .....	161
Figura 74. Confianza - Octavo ciclo Sistemas.....	162
Figura 75. Confianza - Noveno ciclo Sistemas.....	163
Figura 76. Confianza - Décimo ciclo Sistemas .....	163
Figura 77. Confianza - Primer Ciclo Computación .....	165
Figura 78. Confianza - Segundo Ciclo Computación.....	165
Figura 79. Confianza - Tercer Ciclo Computación.....	166
Figura 80. Confianza - Cuarto Ciclo Computación .....	167
Figura 81. Selección de variables para la elaboración del modelo.....	168
Figura 82. Obtención de reglas de asociación .....	169
Figura 83. Reglas obtenidas aplicando filtros (Conf – Lift) .....	169
Figura 84. Reglas obtenidas en Weka con Lift mayor a 1 .....	171
Figura 85. Reglas de asociación en RapidMiner .....	173
Figura 86. Reglas de asociación obtenidas por Python .....	173
Figura 87. Clústeres con sus respectivas instancias.....	174
Figura 88. Clústeres obtenidos por Weka. ....	174
Figura 89. Clúster con total de instancias .....	176
Figura 90. Representación gráfica de los clústeres obtenidos .....	176
Figura 91. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	177
Figura 92. Relación Género - Tipo de Colegio .....	178
Figura 93. Relación Cálculo Diferencial - Género. ....	179
Figura 94. Relación Edad - Fundamentos Informáticos .....	179
Figura 95. Relación Sector de procedencia - Fundamentos Informáticos .....	180
Figura 96. Selección de variables para la elaboración del modelo.....	181
Figura 97. Obtención de reglas de asociación .....	182
Figura 98. Reglas obtenidas aplicando filtros (Conf – Lift) .....	182

Figura 99. Reglas obtenidas en Weka con Lift mayor a 1 .....	183
Figura 100. Reglas de asociación en RapidMiner .....	185
Figura 101. Reglas de asociación obtenidas por Python .....	186
Figura 102. Clúster con sus respectivas instancias.....	186
Figura 103. Clústeres obtenidos por Weka. ....	187
Figura 104. Clúster con total de instancias .....	188
Figura 105. Agrupación de cada Instancia a su respectivo Clúster .....	189
Figura 106. Representación gráfica de los clústeres obtenidos .....	189
Figura 107. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	190
Figura 108. Relación Género - Tipo de Colegio .....	192
Figura 109. Relación Edad – Programación I.....	192
Figura 110. Relación Género - Programación I .....	193
Figura 111. Relación Cálculo Diferencial – Programación I .....	193
Figura 112. Selección de variables para la elaboración del modelo.....	194
Figura 113. Obtención de reglas de asociación .....	195
Figura 114. Reglas obtenidas aplicando filtros (Conf – Lift) .....	195
Figura 115. Reglas obtenidas en Weka con Lift mayor a 1 .....	197
Figura 116. Reglas de asociación en RapidMiner .....	199
Figura 117. Reglas de asociación obtenidas por Python .....	201
Figura 118. clústeres con sus respectivas instancias.....	201
Figura 119. Clúster obtenidos por Weka. ....	201
Figura 120. Clúster con total de instancias .....	203
Figura 121. Agrupación de cada Instancia a su respectivo Clúster.....	203
Figura 122. Representación gráfica de los clústers obtenidos .....	204
Figura 123. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	204
Figura 124. Relación Género - Tipo de Colegio .....	206
Figura 125. Relación Edad – Programación II.....	206
Figura 126. Relación Género – Programación II .....	207
Figura 127. Selección de variables para la elaboración del modelo.....	208
Figura 128. Obtención de reglas de asociación .....	209
Figura 129. Reglas obtenidas aplicando filtros (Conf – Lift) .....	209
Figura 130. Reglas obtenidas en Weka con Lift mayor a 1 .....	211
Figura 131. Reglas de asociación en RapidMiner .....	213
Figura 132. Reglas de asociación obtenidas por Python .....	214

Figura 133. Clúster con sus respectivas instancias.....	214
Figura 134. Clúster obtenidos por Weka. ....	215
Figura 135. Clúster con total de instancias .....	216
Figura 136. Agrupación de cada Instancia a su respectivo Clúster.....	217
Figura 137. Representación gráfica de los clústeres obtenidos .....	217
Figura 138. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	218
Figura 139. Relación Género – Tipo de Colegio .....	220
Figura 140. Relación Edad – Metodología de la Programación .....	220
Figura 141. Relación Género – Metodología de la Programación.....	221
Figura 142. Relación Diseño y Gestión de Base de Datos – Metodología de la Programación.....	221
Figura 143. Selección de variables para la elaboración del modelo.....	222
Figura 144. Obtención de reglas de asociación .....	223
Figura 145. Reglas obtenidas aplicando filtros (Conf – Lift) .....	223
Figura 146. Reglas obtenidas en Weka con Lift mayor a 1 .....	225
Figura 147. Reglas de asociación en RapidMiner .....	227
Figura 148. Reglas de asociación obtenidas por Python .....	228
Figura 149. clústeres con sus respectivas instancias.....	228
Figura 150. Clúster obtenidos por Weka. ....	229
Figura 151. Clúster con total de instancias .....	230
Figura 152. Agrupación de cada Instancia a su respectivo Clúster.....	231
Figura 153. Representación gráfica de los clústeres obtenidos .....	231
Figura 154. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	232
Figura 155. Relación Género – Tipo de Colegio .....	233
Figura 156. Relación Edad – Metodología de la Programación .....	234
Figura 157. Relación Género – Metodología de la Programación.....	234
Figura 158. Relación Diseño y Gestión de Base de Datos – Metodología de la Programación.....	235
Figura 159. Selección de variables para la elaboración del modelo.....	236
Figura 160. Obtención de reglas de asociación .....	237
Figura 161. Reglas obtenidas aplicando filtros (Conf – Lift) .....	237
Figura 162. Reglas obtenidas en Weka con Lift mayor a 1 .....	239
Figura 163. Reglas de asociación en RapidMiner.....	241
Figura 164. Reglas de asociación obtenidas por Python .....	242

Figura 165. clústeres con sus respectivas instancias.....	242
Figura 166. Clúster obtenidos por Weka. ....	243
Figura 167. Clúster con total de instancias .....	244
Figura 168. Agrupación de cada Instancia a su respectivo Clúster.....	245
Figura 169. Representación gráfica de los clústers obtenidos .....	245
Figura 170. Clúster obtenidos por Auto model de RapidMiner.....	246
Figura 171. Relación Género – Tipo de Colegio .....	247
Figura 172. Relación Edad – Simulación .....	248
Figura 173. Relación Género – Simulación.....	248
Figura 174. Relación Lenguaje Ensamblador - Simulación.....	249
Figura 175. Selección de variables para la elaboración del modelo.....	250
Figura 176. Obtención de reglas de asociación .....	251
Figura 177. Reglas obtenidas aplicando filtros (Conf – Lift) .....	251
Figura 178. Reglas obtenidas en Weka con Lift mayor a 1 .....	253
Figura 179. Reglas de asociación en RapidMiner.....	254
Figura 180. Reglas de asociación obtenidas por Python .....	255
Figura 181. clústeres con sus respectivas instancias.....	255
Figura 182. Clúster obtenidos por Weka. ....	256
Figura 183. Clúster con total de instancias .....	257
Figura 184. Agrupación de cada Instancia a su respectivo Clúster.....	258
Figura 185. Representación gráfica de los clústers obtenidos .....	258
Figura 186. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	259
Figura 187. Relación Género – Tipo de Colegio .....	261
Figura 188. Relación Edad – Ecuaciones Diferenciales.....	261
Figura 189. Relación Género – Ecuaciones Diferenciales .....	262
Figura 190. Relación Sistemas Operativos - Ecuaciones Diferenciales .....	262
Figura 191. Selección de variables para la elaboración del modelo.....	263
Figura 192. Obtención de reglas de asociación .....	264
Figura 193. Reglas obtenidas aplicando filtros (Conf – Lift) .....	265
Figura 194. Reglas obtenidas en Weka con Lift mayor a 1 .....	267
Figura 195. Reglas de asociación en RapidMiner.....	270
Figura 196. Reglas de asociación obtenidas por Python .....	271
Figura 197. clústeres con sus respectivas instancias.....	271
Figura 198. Clúster obtenidos por Weka. ....	271

Figura 199. Clústers con total de instancias.....	273
Figura 200. Agrupación de cada Instancia a su respectivo Clúster.....	273
Figura 201. Representación gráfica de los clústers obtenidos .....	274
Figura 202. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	274
Figura 203. Relación Género – Tipo de Colegio .....	276
Figura 204. Relación Edad – Análisis Numérico .....	276
Figura 205. Relación Género – Análisis Numérico.....	277
Figura 206. Relación Gestión de Redes – Investigación de Operaciones.....	277
Figura 207. Selección de variables para la elaboración del modelo.....	278
Figura 208. Obtención de reglas de asociación .....	279
Figura 209. Reglas obtenidas aplicando filtros (Conf – Lift) .....	279
Figura 210. Reglas obtenidas en Weka con Lift mayor a 1 .....	281
Figura 211. Reglas de asociación en RapidMiner.....	284
Figura 212. Reglas de asociación obtenidas por Python .....	285
Figura 213. Clúster con sus respectivas instancias.....	285
Figura 214. Clústeres obtenidos por Weka. ....	285
Figura 215. Clúster con total de instancias .....	287
Figura 216. Agrupación de cada Instancia a su respectivo Clúster.....	287
Figura 217. Representación gráfica de los clústeres obtenidos .....	288
Figura 218. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	288
Figura 219. Relación Género – Tipo de Colegio .....	290
Figura 220. Relación Edad – Ingeniería de Software II .....	290
Figura 221. Relación Género – Ingeniería de Software II .....	291
Figura 222. Relación Modelamiento Matemático – Ingeniería de Software II	291
Figura 223. Selección de variables para la elaboración del modelo.....	292
Figura 224. Obtención de reglas de asociación .....	293
Figura 225. Reglas obtenidas aplicando filtros (Conf – Lift) .....	293
Figura 226. Reglas obtenidas en Weka con Lift mayor a 1 .....	295
Figura 227. Reglas de asociación en RapidMiner.....	297
Figura 228. Reglas de asociación obtenidas por Python .....	298
Figura 229. Clúster con sus respectivas instancias.....	298
Figura 230. Clústeres obtenidos por Weka. ....	298
Figura 231. Clústers con total de instancias.....	300
Figura 232. Clúster con total de instancias .....	300

Figura 233. Representación gráfica de los clústers obtenidos .....	301
Figura 234. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	301
Figura 235. Relación Género – Tipo de Colegio .....	303
Figura 236. Relación Edad – Control Automatizado Asistido por computadores .....	303
Figura 237. Relación Género – Control Automatizado Asistido por computadores .....	304
Figura 238. Relación Compiladores – Control Automatizado Asistido por computadores.....	305
Figura 239. Selección de variables para la elaboración del modelo.....	306
Figura 240. Obtención de reglas de asociación .....	307
Figura 241. Reglas obtenidas aplicando filtros (Conf – Lift) .....	307
Figura 242. Reglas obtenidas en Weka con Lift mayor a 1 .....	310
Figura 243. Reglas de asociación en RapidMiner .....	313
Figura 244. Reglas de asociación obtenidas por Python .....	314
Figura 245. Clúster con sus respectivas instancias.....	314
Figura 246. Clústeres obtenidos por Weka. ....	314
Figura 247. Clúster con total de instancias .....	316
Figura 248. Agrupación de cada Instancia a su respectivo Clúster.....	316
Figura 249. Representación gráfica de los clústeres obtenidos .....	317
Figura 250. Clústeres obtenidos por Auto Model dentro de Rapidminer .....	317
Figura 251. Relación Género – Tipo de Colegio .....	319
Figura 252. Relación Edad – Ingeniería de Software II .....	319
Figura 253. Relación Género – Ingeniería de Software II .....	320
Figura 254. Relación Modelamiento Matemático – Ingeniería de Software II .....	321
Figura 255. Selección de variables para la elaboración del modelo.....	322
Figura 256. Obtención de reglas de asociación .....	323
Figura 257. Reglas obtenidas aplicando filtros (Conf – Lift) .....	323
Figura 258. Reglas obtenidas en Weka con Lift mayor a 1 .....	326
Figura 259. Reglas de asociación en RapidMiner .....	329
Figura 260. Reglas de asociación obtenidas por Python .....	329
Figura 261. Clústers con sus respectivas instancias.....	330
Figura 262. Clústers obtenidos por Weka. ....	330
Figura 263. Clúster con total de instancias .....	332

Figura 264. Agrupación de cada Instancia a su respectivo Clúster.....	332
Figura 265. Representación gráfica de los clústeres obtenidos.....	333
Figura 266. Clústeres obtenidos por Auto Model dentro de Rapidminer.....	333
Figura 267. Relación Género – Tipo de Colegio.....	335
Figura 268. Relación Edad – Ingeniería de Software II.....	335
Figura 269. Relación Género – Ingeniería de Software II.....	336
Figura 270. Relación Modelamiento Matemático – Ingeniería de Software II.....	336
Figura 271. Selección de variables para la elaboración del modelo.....	337
Figura 272. Obtención de reglas de asociación.....	338
Figura 273. Reglas obtenidas aplicando filtros (Conf – Lift).....	338
Figura 274. Reglas obtenidas en Weka con Lift mayor a 1.....	341
Figura 275. Reglas de asociación en RapidMiner.....	344
Figura 276. Reglas de asociación obtenidas por Python.....	345
Figura 277. Clústers con sus respectivas instancias.....	345
Figura 278. Clústers obtenidos por Weka.....	345
Figura 279. Clúster con total de instancias.....	347
Figura 280. Agrupación de cada Instancia a su respectivo Clúster.....	348
Figura 281. Representación gráfica de los clústers obtenidos.....	348
Figura 282. Relación Género – Tipo de Colegio.....	349
Figura 283. Relación Edad – Ingeniería de Software II.....	349
Figura 284. Relación Género – Ingeniería de Software II.....	350
Figura 285. Relación Modelamiento Matemático – Ingeniería de Software II.....	351
Figura 286. Selección de variables para la elaboración del modelo.....	352
Figura 287. Obtención de reglas de asociación.....	353
Figura 288. Reglas obtenidas aplicando filtros (Conf – Lift).....	353
Figura 289. Reglas obtenidas en Weka con Lift mayor a 1.....	355
Figura 290. Reglas de asociación en RapidMiner.....	359
Figura 291. Reglas de asociación obtenidas por Python.....	359
Figura 292. Clústers con sus respectivas instancias.....	360
Figura 293. Clústers obtenidos por Weka.....	360
Figura 294. Clúster con total de instancias.....	362
Figura 295. Agrupación de cada Instancia a su respectivo Clúster.....	362
Figura 296. Representación gráfica de los clústeres obtenidos.....	363
Figura 297. Relación Género – Tipo de Colegio.....	364

Figura 298. Relación Edad – Ingeniería de Software II.....	364
Figura 299. Relación Género – Ingeniería de Software II .....	365
Figura 300. Relación Modelamiento Matemático – Ingeniería de Software II	365



## ÍNDICE DE TABLAS

Tabla I. Plan de acción.....	49
Tabla II. Recursos humanos.....	62
Tabla III. Recursos materiales.....	63
Tabla IV. Recursos técnicos.....	63
Tabla V. Costo aproximado del proyecto .....	63
Tabla VI. Cadenas De Búsqueda Para La RSL.....	84
Tabla VII. Cadenas de búsqueda relacionadas a las herramientas para la MD. .....	85
Tabla VIII. Cadena de búsqueda para la metodología a usar en el desarrollo de MD.....	86
Tabla IX. Trabajos relacionados obtenidos de la RSL.....	88
Tabla X. Trabajos relacionados con la herramienta a utilizar .....	98
Tabla XI. Trabajos relacionados con la metodología a utilizar .....	105
Tabla XII. Detalle de cada variable presente en el conjunto de datos inicial..	115
Tabla XIII. Variables dentro del conjunto de datos. ....	134
Tabla XIV. Variables para la minería de datos .....	138
Tabla XV. Diccionario de variables sistemas/computación .....	141
Tabla XVI. Codificación para las provincias de nacimiento sistemas/computación .....	141

## 1. TITULO

# **Minería de Datos Educativa aplicada al descubrimiento de patrones en registros académicos de estudiantes de Ingeniería con asignaturas relacionadas a las matemáticas**

## 2. RESUMEN

A nivel Mundial, la deserción académica es un fenómeno que subsiste en todas las Instituciones de Educación Superior. En Ecuador, según reporte presentado por la UNESCO existe un 40% de abandono estudiantil universitario; en algunas Instituciones se ha demostrado que existen diversas causas de deserción, como factores sociales, académicos, demográficos, económicos, pedagógicos, entre otros. Pese a que se ha tratado, a través de diferentes técnicas, minorar los porcentajes de abandono académico, se ha logrado identificar que estos índices han aumentado debido al programa de admisión a la Universidad implementada por el Gobierno. Añadiendo que, existe bajo interés en los estudiantes con asignaturas relacionadas a las Matemáticas en Carreras de Ingeniería ya que, de acuerdo con los resultados obtenidos por la Prueba “Ser Bachiller”, se alcanza únicamente un nivel de “Elemental” en dicha área, añadiendo al sector rural como el más afectado.

En virtud de lo anteriormente expuesto, el objetivo principal del presente Trabajo de Titulación es aplicar técnicas de Minería de Datos Educativa para identificar los patrones en los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas, en la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, mismo que se lo realizó con base a las fases del Metodología CRISP-DM, las cuales fueron desarrolladas a través de Excel, Google Colab y mediante el uso de librerías de Python como son Scikit-Learn y Pandas. Los registros académicos se proporcionaron desde la Dirección de Telecomunicaciones e Información UTI, mismos que fueron preparados para el tema de estudio a través de Python, durante este proceso, se empleó la técnica de K-vecino más cercano para el relleno de algunos datos, obteniendo como salida un modelo con el cual aplicados los diferentes algoritmos, uno por cada ciclo, se obtienen los resultados del presente TT. Para el descubrimiento de patrones se emplearon técnicas descriptivas mismas que se ejecutaron a través de: Reglas de Asociación (A-priori) y Clústering (K-means), implementadas en tres herramientas, Weka, RapidMiner y Python. Finalmente, a través de su ejecución y con la información obtenida por cada ambiente se identifican a factores como no trabajar, del sector urbano, de género masculino, solteros como los más predominantes entre la agrupación de variables sociales, demográficas y académicas y a partir de ellas, se propone un plan de acción a nivel general, con estrategias que sirvan de apoyo a las autoridades correspondientes en la toma de decisiones, con el fin de minorar

los índices de deserción estudiantil en la Carrera de Ingeniería en Sistemas/Computación.

**Palabras Claves:** Registro académico, Asignatura de matemáticas, Ingeniería, Minería de Datos Educativa, Crisp-DM, Técnicas Descriptivas, Weka, RapidMiner, Python.

## ABSTRACT

Worldwide, academic desertion is a phenomenon that persists in all higher education institutions. In Ecuador, according to a report presented by UNESCO, there is a 40% university student dropout rate; in some institutions it has been shown that there are various causes of dropout, such as social, academic, demographic, economic, pedagogical factors, among others. Although it has been tried, through different techniques, to reduce the dropout rates, it has been identified that these rates have increased due to the University admission program implemented by the Government. In addition, there is low interest in students with subjects related to Mathematics in Engineering careers since, according to the results obtained by the "Ser Bachiller" test, only an "Elementary" level is reached in this area, adding the rural sector as the most affected.

By virtue of the above, the main objective of this Degree Project is to apply Educational Data Mining techniques to identify patterns in the academic records of students with subjects related to mathematics, in the Systems/Computer Engineering Career of the National University of Loja, which was carried out based on the phases of the CRISP-DM Methodology, which were developed through Excel, Google Colab and through the use of Python libraries such as Scikit-Learn and Pandas. The academic records were provided by the Directorate of Telecommunications and Information UTI, which were prepared for the topic of study through Python. During this process, the K-nearest neighbor technique was used to fill in some data, obtaining as output a model with which the different algorithms were applied, one for each cycle, to obtain the results of this TT. For the discovery of patterns, descriptive techniques were used, which were executed through: Association Rules (A-priori) and Clustering (K-means), implemented in three tools, Weka, RapidMiner and Python. Finally, through its execution and with the information obtained by each environment, factors such as not working, urban sector, male gender, single, are identified as the most predominant among the grouping of social, demographic and academic variables and from them, an action plan is proposed at a general level, with strategies that support the corresponding authorities in decision making, in order to reduce student desertion rates in the Systems/Computer Engineering Career.

**Keywords:** Academic Record, Mathematics Subject, Engineering, Educational Data Mining, Crisp-DM, Descriptive Techniques, Weka, RapidMiner, Python.

### 3. INTRODUCCION

La deserción en la educación superior es una problemática que afecta a todas las instituciones de educación superior a nivel mundial, cada una de ellas asociadas a causas sociales, académicos, psicológicas o demográficas. La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura UNESCO informa que el 40 % de estudiantes abandonan su Carrera Universitaria, donde se identifica que la mayor tasa de deserción se da en los primeros ciclos universitarios [1]. Algunos trabajos [1]–[6] hacen referencia a los factores personales, académicos y socioeconómicos como los más comunes para que los estudiantes abandonen la Carrera. Por otro lado, Fernández [7] determina que existen factores pedagógicos (metodología de enseñanza-aprendizaje) que se involucran en el abandono de los estudiantes.

Así mismo, existen algunos estudios que analizan la deserción en Universidades vinculadas a Carreras de Ingenierías y cursos de Matemática [8] . La Matemática es una asignatura básica que, dada su versatilidad se manifiesta no solo como herramienta en otras disciplinas científicas, sino también en múltiples actividades profesionales como es el caso de las Ingenierías [9].

Según datos presentados por el Instituto Nacional de Evaluación Educativa (INEVAL) en el programa para la Evaluación de Estudiantes para el desarrollo (PISA-D), desarrollada en el año 2017, el nivel mínimo de habilidades en Matemática alcanza únicamente el 29% reflejándose así una alta deficiencia en la formación de los estudiantes desde etapas iniciales [10]. De igual forma, en los resultados de los exámenes obtenidos por estudiantes de secundaria denominado “Ser Bachiller”, en el año 2017 - 2018, se ha demostrado que el grado de “Insuficiencia” y “Elemental”, predominan en el campo de las matemáticas con 16.9% y 48,2% respectivamente [11]; en el año 2018 – 2019, el promedio en dominio matemático es de 7,98 puntos sobre 10, es decir, se encuentra en un nivel de logro “Elemental”, considerando que el sector rural refleja un índice de deficiencia matemática de 0,17 en comparación al sector urbano. Además, se ha demostrado que existen mejores resultados en instituciones particulares que en instituciones públicas (fiscales) [12]. Esto presentaría un inconveniente en las Carreras del futuro debido al bajo potencial de conocimiento que se presenta en los aspirantes que intentan ingresar a la Universidad.

En la ciudad de Loja, dentro de las Instituciones a nivel Superior, se ha demostrado que se están dejando de lado las Carreras de los campos de conocimiento vinculados con las Carreras STEAM Science, Technology, Engineering and Mathematics (Ciencia, Tecnología, Ingeniería y Matemáticas) tales como: “Ingeniería, Industria y Construcción”, “Tecnologías de la Información y Comunicación”, “Ciencias Naturales, Matemáticas y Estadística” [13].

En virtud de lo anteriormente expuesto, resulta importante determinar cuáles son los factores predominantes dentro de los registros académicos de los estudiantes de la Carrera de Ingeniería en Sistemas/Computación, relacionándolos con las asignaturas que, de acuerdo con el código de la UNESCO pertenecen a las matemáticas. Con base a ello se pretende conocer dicha información y proponer un plan de acción que sirva de apoyo para la toma de decisiones a las autoridades correspondientes y a su vez mitigar la deserción académica universitaria.

Los objetivos específicos en los que se basa el desarrollo del presente TT son tres: Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería. Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa. Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación; con el propósito de llevar a cabo el objetivo general: Aplicar técnicas de Minería de Datos Educativa para identificar patrones en los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas en la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja; ya que su ejecución conduce a dar respuesta a la pregunta de investigación: ¿El uso de técnicas de minería de datos educativa permitirá identificar los patrones que existen en los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas en la Carrera de Ingeniería en Sistemas/Computación?.

Para el desarrollo del TT, se inició con el cumplimiento del primer objetivo específico a través de la ejecución de la Revisión Sistemática de Literatura RSL, misma que proporcionó un sustento teórico al TT, así como también aplicada para dar respuesta a las cinco preguntas de investigación, proporcionando las bases necesarias y concretas de los conocimientos, técnicas y herramientas para llevar a cabo el TT.

De la misma manera, permitió identificar a trabajos en donde describen a las causas de abandono universitario. El segundo objetivo comprendió la búsqueda de los patrones en los registros académicos de los estudiantes de Sistemas/Computación para ello se utilizó la metodología CRISP-DM, la misma que permitió llevar un control y monitoreo de todo el TT, de la misma manera, se utiliza los resultados de la RSL para determinar la herramienta con la que se va a trabajar y ejecutar las técnicas de la minería de datos educativa. Finalmente, para el cumplimiento del tercer objetivo se inició con elaboración de la propuesta de un plan de acción basado en los resultados obtenidos por el segundo objetivo, mismo que está dirigido para la toma de decisiones de las autoridades correspondientes con el propósito de mitigar la deserción académica en la Carrera de Ingeniería en Sistemas/Computación.

El Proceso investigativo seguido para alcanzar los objetivos, con base a los lineamientos de la Universidad Nacional de Loja, inicia con una Revisión de Literatura, donde se detallan los conceptos concisos referentes al tema de estudio como la deserción universitaria, matemáticas, Inteligencia Artificial, Minería de Datos Educativa, Técnicas de Minería de Datos y Metodología de Minería de Datos; en la sección de materiales y métodos se detalla el contexto dentro del que se enmarca el TT, el proceso y los recursos científicos, técnicos y éticos que este utilizaron durante el desarrollo del trabajo investigativo. En la sección de resultados se muestran los principales hallazgos producto de cada una de las fases consideradas para dar cumplimiento a los objetivos específicos. La sección de discusión describe el aporte del investigador con relación a los resultados obtenidos por las actividades desarrolladas durante el proceso investigativo de cada una de las fases, asimismo, se muestra la valoración técnica, económica y ambiental que genera el desarrollo y ejecución del trabajo investigativo. Finalmente, se detallan las conclusiones y recomendaciones determinadas por el investigador una vez realizado el TT; así como también se exponen las limitaciones que no se abarcan el trabajo investigativo por temas de tiempo y recursos, finalmente se muestran los trabajos a futuro que se pueden obtener con base al presente trabajo investigativo.



## **4. REVISIÓN DE LITERATURA**

La presente sección brinda las bases teóricas que sustentan al TT, así como también los conceptos relevantes que permiten dar un mejor conocimiento acerca del tema. Se inicia con una reseña histórica y las definiciones acerca de la deserción universitaria expuestas en la sección 4.1. A continuación en la sección 4.2 se presentan los conceptos relevantes sobre las matemáticas, seguidamente en la sección 4.3 manifiesta a la Inteligencia Artificial (IA), así también se presenta la sección 4.4 a uno de los campos de la IA como es la Minería de Datos; en la sección 4.5 se describen las Técnicas de Minería de Datos, seguidamente la sección 4.6 detalla algunos los conceptos acerca de las herramientas más utilizadas para la Minería de Datos para finalmente, en sección 4.7 se define a la Metodología CRISP-DM para realizar la Minería de Datos.

### **4.1. Deserción Universitaria**

La deserción estudiantil universitaria es uno de los problemas más importantes con que se enfrenta el sistema de educación superior. Muchas de las autoridades dentro de las instituciones han venido presentando su preocupación ante este inconveniente debido a que el número de estudiantes que logran culminar sus estudios no es el esperado, puesto que muchos logran ingresar y en su mayoría desertan en los primeros ciclos de su Carrera universitaria. Es por ello que nace conocer ¿Por qué el rendimiento académico de los estudiantes recién ingresados a la Universidad es menor al de los estudios secundarios (colegio), sabiendo que eran buenos alumnos en la misma?, ¿es responsabilidad de la Universidad que la mayoría de los estudiantes recién ingresados abandonen su Carrera?

Existe una brecha académica entre lo que recibe el alumno en la escuela y lo que exige la universidad, este problema se ve reflejado a nivel mundial. El porcentaje de alumnos desaprobados en una o más asignaturas del primer ciclo es elevado, lo cual lleva a que varios de ellos se matriculen nuevamente en las mismas; y a pesar de esto se vuelve a repetir el proceso y esto provoca el abandono de la universidad, provocando en el país un bajo número de profesionales [14].

#### **4.1.1. Definición**

Es considerada como uno de los puntos más críticos dentro de la educación superior. Según Calderón [4] considera a la deserción como “el proceso de abandono, voluntario o forzoso de la Carrera en la que se matricula un estudiante, por la influencia positiva o negativa de circunstancias internas”.

Por otra parte, Perassi [15] hace referencia a dejar inconcluso el camino planteado por el sistema educativo y es el punto culminante de la cadena del fracaso escolar ya que, es lo más frecuente que antes de desertar haya repetido año o reprobado materias, alargando su trayecto escolar y debilitando su autoestima.

La deserción dentro de la Universidad Nacional de Loja, al igual que en otras instituciones de educación superior, presenta gran preocupación a las autoridades debido a que dentro de las Carreras de Ingeniería el número de abandono estudiantil es mucho más elevado en comparación a otras Carreras pertenecientes a la misma; es por ello que dentro del TT se busca en base a un plan de acción presentar algunas alternativas que permitan a las autoridades mejorar el proceso de toma de decisiones que ayuden a mitigar la deserción universitaria.

#### **4.1.2. Factores de la deserción Universitaria**

La deserción estudiantil dentro de las Universidades es determinada por distintos factores que propician a esta, como los externos, estos pueden ser: Los Factores económicos, dentro de ellos tenemos los bajos ingresos familiares, desempleo cabeza de familia, incompatibilidad entre trabajo y estudio, es decir la situación laboral del estudiante, falta de apoyo familiar, situación económica precaria del estudiante y en otro casos la influencia del nivel educativo de los padres; de manera similar, dentro de los factores académicos se pueden dar por el bajo rendimiento, elección inadecuada de la carrera, repetición de las asignaturas, colegio de procedencia del estudiante (considerado un factor importante puesto que según [12] los colegios privados muestran mejores resultados de aprendizaje en comparación a los fiscomisionales y públicos), métodos de estudio, es decir la metodología que utiliza el docente para llegar a los estudiantes, la

carga académica que se presenta dentro de la carrera, la insatisfacción con el programa universitario y en otros casos inconformidad con las calificaciones obtenidas con el examen de admisión [1], [14].

Añadiendo a lo anteriormente expuesto, la familia otro factor importante que influye en la deserción universitaria, existen diferentes niveles o estratos, como son altos, medio y bajos. Las familias de estratos bajos, tienen como única preocupación conseguir dinero para subsistir, dejando a un lado la educación, además, en muchas ocasiones vienen de familias con un solo padre y varios hijos, situación que hace casi imposible sostenerlos en la escuela sin importar que sea gratuita. Por su parte, las familias de clase media ya con una mejor posición se preocupan por tener una buena educación y un mejor futuro; mientras que, los de estratos altos velan porque sus hijos tengan la mejor educación sin importan la atención que ellos les brinden a sus hijos, los estudiantes de estrato social bajo se ven afectados [16]. Ante esto se puede contrastar que todos los factores antes mencionados (Ver Figura 1) implican de manera directa o indirectamente a que el estudiante se vea obligado a desertar sus estudios.

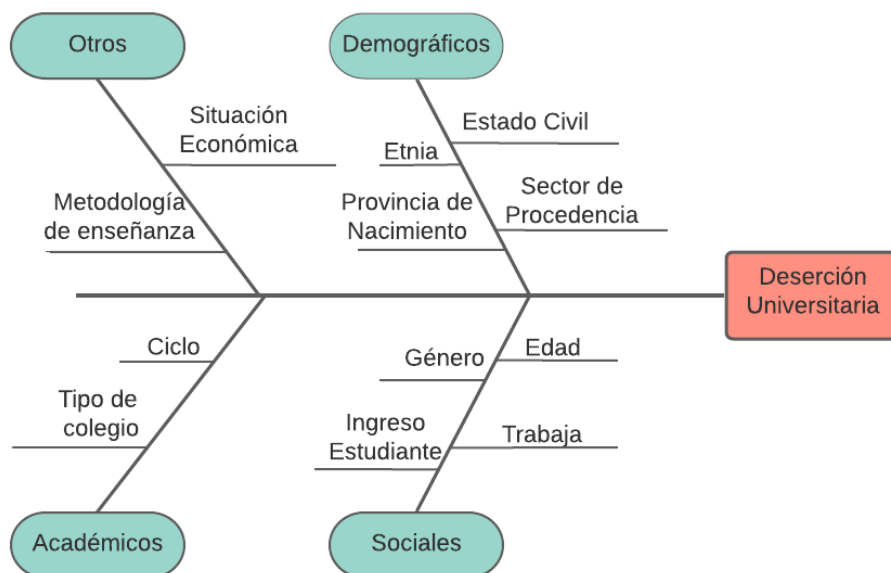


Figura 1. Factores de Deserción Universitaria

## **4.2. Las Matemáticas**

Según el Diccionario de la Real Academia Española (RAE)<sup>1</sup>, define a la matemática como: “Ciencia deductiva que estudia las propiedades de los entes abstractos, como números, figuras geométricas o símbolos, y sus relaciones”.

La Matemática, como una de las manifestaciones de la mente humana, refleja la voluntad activa, la razón contemplativa y el deseo de perfección estética; sus elementos básicos son lógica e intuición, deducción y construcción, análisis y síntesis, generalidad y particularidad. La exclusión de algunos de ellos y en particular la imposición del rigor como valor fundamental y casi único de la Matemática, ha tenido serias consecuencias en el ámbito escolar. A veces se exige a los estudiantes un rigor que no responde a las necesidades de los problemas que se tratan, de modo que puede hacer que el razonamiento matemático aparente ser un juego absurdo donde se pretende demostrar lo evidente. Desde el punto de vista del estudiante a veces se presenta como una circunstancia extraña que muchos teoremas son más evidentes que los axiomas y premisas usados para demostrarlos. Esto desconcierta al estudiante que puede llegar a pensar que las Matemáticas son un juego inútil. El rigor se ha utilizado para dar un aspecto de profundidad a las Matemáticas elementales. Ello ha propiciado un fracaso escolar grave y por tanto un fracaso pedagógico, y no sólo en los niveles elementales y medios, sino también en los universitarios [17].

### **4.2.1. Influencia de las Matemáticas en los estudiantes universitarios**

La Matemática está presente en muchas disciplinas científicas, como es el caso de las Carreras técnicas o Ingenierías, donde, aun siendo necesarias muchas habilidades como la comunicación, el trabajo colaborativo, entre otros, las destrezas técnicas en Matemáticas siguen constituyendo la base de la formación de los Ingenieros. Esto supone que los estudiantes que deciden matricularse en dichas Carreras poseen un alto interés y competitividad hacia las matemáticas. No obstante, en

---

<sup>1</sup> <https://dle.rae.es/>

investigaciones se revelan elevadas cifras de suspensos, desinterés, frustración, angustia y temor hacia esta asignatura a nivel personal y entre el alumnado [9].

A más de los factores económicos, sociales, familiares, académicos (centrándonos en la enseñanza), la falta o deficiente interacción entre docente y estudiante, se convierte en una causa para provocar la deserción. La entrega de una experiencia de aprendizaje deficiente, poco personalizada, puede llegar a desmotivar a los estudiantes, convirtiéndose en una experiencia de insatisfacción personal [7].

Aunado a esto, la manera en cómo el docente imparte sus enseñanzas hacia los alumnos es de gran importancia para evitar la deserción, se dice que una actitud positiva por parte del docente ayuda a los alumnos a tomar la iniciativa, centrándose en el descubrimiento y provocando en los alumnos el gusto y confianza con las matemáticas. Por el contrario, un profesor con actitudes negativas utiliza métodos de enseñanza que fomentan sentimientos semejantes a los suyos como la inseguridad, desmotivación, la ansiedad, la misma falta de conocimiento y por ende el disgusto con la asignatura [9].

### **4.3. Inteligencia Artificial**

Desde sus orígenes la inteligencia artificial (IA) ha demostrado ser un factor importante para el ser humano, destacándose principalmente en brindar ayuda en la toma de decisiones, ya que a través de la IA se pueden llevar a cabo acciones que guíen a muchas personas a realizar un cambio.

#### **4.3.1. Definición de la Inteligencia Artificial**

Según [18] se define a la inteligencia artificial como “la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano”.

#### **4.3.2. Inteligencia Artificial en la Educación**

La IA tiene un fuerte potencial para acelerar el proceso de realización y desarrollo de los objetivos globales en torno a la educación mediante la reducción de las dificultades de acceso al aprendizaje, la automatización de los procesos de gestión y la optimización de los métodos que permiten

mejorar los resultados en el aprendizaje, no obstante, la integración de la IA a los entornos educativos en determinados ambientes puede provocar tiempo tardío debido a las políticas y procesos administrativos de cada institución, sin embargo, en el actual contexto global de la revolución tecnológica existen cualidades humanas que todavía no pueden ser reproducidas por la inteligencia artificial como la creatividad, la capacidad de reproducir nuevas ideas o la capacidad de improvisar y evolucionar constantemente con el tiempo estas limitantes que poco a poco van siendo superadas para alcanzar un desarrollo más óptimo que permita ir más allá de la Industria 4.0 [19].

De acuerdo a [20], en el ámbito de la formación universitaria, la aparición de las nuevas tecnologías digitales está imponiendo a los docentes cambios pedagógicos y metodológicos muy radicales.

#### **4.4. Minería de Datos**

Desde hace tiempos remotos la minería ha sido parte de nuestra vida, desde los años setenta varios estadísticos hacían usos a términos como data mining, data fishing o data archaeology con el propósito de encontrar relaciones sin la necesidad de una previa hipótesis en bases de datos con ruido. Esta tecnología ha presentado un punto de encuentro importante para el ámbito académico y el mundo de los negocios puesto que se hace uso de los problemas que se presentan en la vida cotidiana para en base a esta presentar información oculta dentro de los datos [13].

##### **1.1.1. Definición**

La DM es considerada como un mecanismo de explotación que consiste en la búsqueda y el descubrimiento de información valiosa dentro de grandes volúmenes de datos (Bases de datos), esto es, un proceso de análisis que trabaja a un nivel de conocimiento con el fin de descubrir patrones, relaciones o incluso excepciones útiles asimismo la generalización de modelos predictivos que proporcionen patrones de conocimiento de alto impacto y puedan llegar a la toma de decisiones [21].

### 1.1.2. Minería de Datos Educativa

Las técnicas de minería de datos pueden descubrir información útil que se puede usar en la evaluación formativa para ayudar a los educadores establecer una base pedagógica para las decisiones al momento modificar un entorno o enfoque de enseñanza. La aplicación de la minería de datos en los sistemas educativos es un problema de ciclo activo de formación de hipótesis, pruebas y refinamiento (Ver Figura 2). El conocimiento minado debe entrar en el ciclo del sistema y guía, facilitar y mejorar el aprendizaje como un todo. No solo convertir los datos en conocimiento, sino también conocimiento extraído para la toma de decisiones. Como podemos ver en la Figura 2, los educadores y académicos son los responsables de diseñar, planificar, construir y del mantenimiento de los sistemas educativos. Los estudiantes buscan interactuar con ellos. A partir de toda la información disponible sobre cursos, estudiantes, uso e interacción, se pueden aplicar técnicas de minería de datos para descubrir conocimiento útil que ayuda a mejorar el programa de aprendizaje impuesto. El conocimiento descubierto puede ser utilizado no solo por educadores, también por usuarios propios o estudiantes. Entonces la aplicación de minería de datos en sistemas educativos puede ser orientado a diferentes actores con cada punto particular [22].

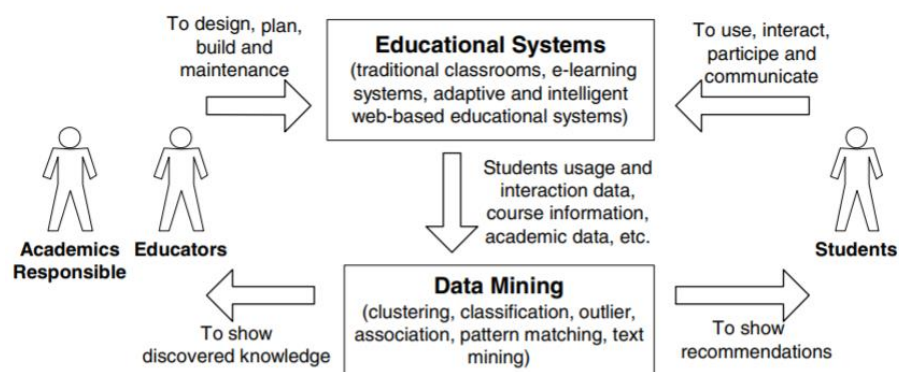


Figura 2. El ciclo de aplicación de minería de datos en sistemas educativos  
La minería de datos se puede aplicar a los datos que provienen de dos tipos de sistemas educativos: aula tradicional y educación a

distancia. Es necesario tratar por separado con la aplicación de técnicas de minería de datos en cada tipo debido a que tienen diferentes fuentes de datos y objetivos.

Los entornos de aula tradicionales son los sistemas educativos más utilizados. Se basa en el contacto cara a cara entre educadores y estudiantes organizados a través de profesores. Han sido criticados porque fomentan el aprendizaje pasivo, ignoran las diferencias individuales y las necesidades de los estudiantes, así como también la falta de atención a la resolución de problemas, pensamiento crítico u otras habilidades de pensamiento de orden superior. En las aulas convencionales, los educadores intentan mejorar las instrucciones al monitorear los procesos de aprendizaje de los estudiantes y analizar su desempeño mediante registros en papel y observación. También pueden usar información sobre asistencia de estudiantes, información de cursos, metas curriculares y datos de planes individualizados. Y la institución educativa tiene muchas fuentes diversas y variadas de información en comparación a las bases de datos tradicionales (con información de un estudiante, información del educador, información de clase y horario, etc.), información en línea (páginas web en línea y páginas de contenido del curso), bases de datos multimedia, etc.[23].

#### **4.5. Técnicas de Minería de Datos**

Una técnica constituye el enfoque conceptual para extraer la información de los datos que se desean tratar, misma que puede ser implementada por diversos algoritmos [24]. Las técnicas de minería de datos crean modelos que son predictivos y/o descriptivos. Un modelo predictivo responde preguntas sobre datos futuros. Un modelo descriptivo proporciona información sobre las relaciones entre los datos y sus características [25].

La Figura 3 muestra las técnicas de minería de datos y los algoritmos que se pueden implementar considerando la data a tratar.



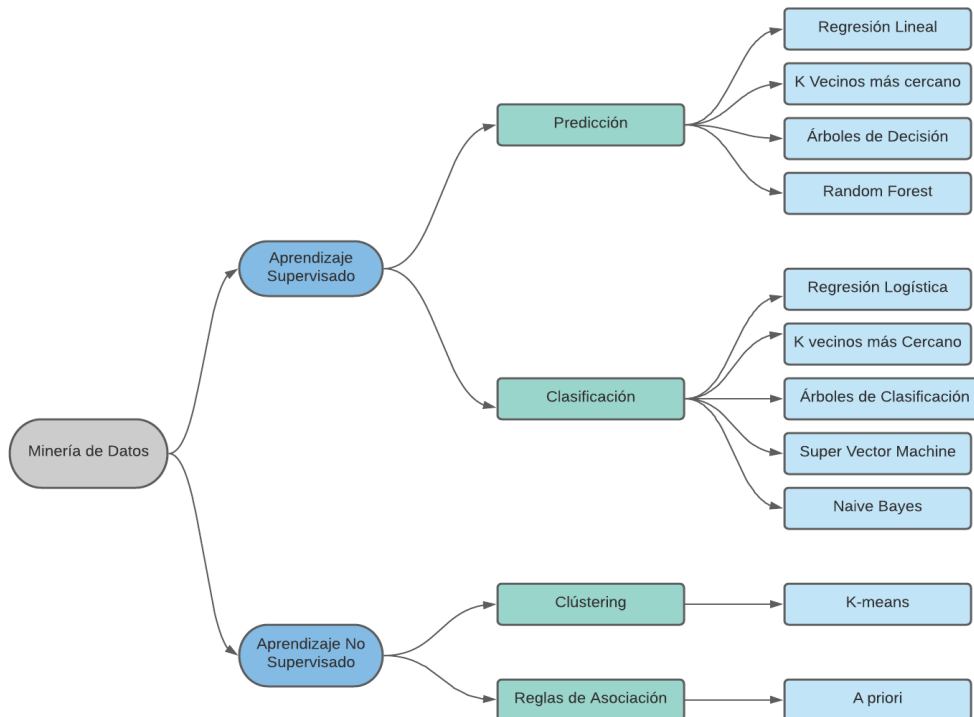


Figura 3. Técnicas de Minería de Datos

#### 4.5.1. Técnicas Supervisadas o Predictivas

Los algoritmos de aprendizaje supervisado o Minería de datos predictiva, obtienen a partir de los datos un modelo que relaciona el valor de un atributo llamado etiqueta y los valores de otros atributos (descriptivos). Si la etiqueta es discreta el modelo será de clasificación mientras que si es continua tendremos un modelo de regresión. El modelo obtenido se utilizará posteriormente para realizar predicciones con datos no etiquetados [26].

#### 4.5.2. Técnicas No Supervisadas o Descriptivas

Las técnicas de minería de datos no supervisadas, también conocidas con el nombre de técnicas de descubrimiento del conocimiento, se utilizan para la detección de patrones ocultos en bases de datos de gran tamaño. Dichos patrones representan por sí mismos información útil que puede ser utilizada directamente en la toma de decisiones [26].

#### 4.6. Herramientas Minería de Datos

Existen un sin número de herramientas que se emplean para realizar la minería de datos, estas proporcionan al usuario un entorno para que se

ejecuten los diferentes algoritmos. De acuerdo a resultados emitidos por KDnuggets en el año 2014 (Ver Figura 4), se han obtenidos los siguientes porcentajes de acuerdo a cada herramienta considerando que existen más herramientas que se identifican por la encuesta. Así mismo, en la Figura 27, se muestran resultados obtenidos de documentos relacionados con el tema de estudio.

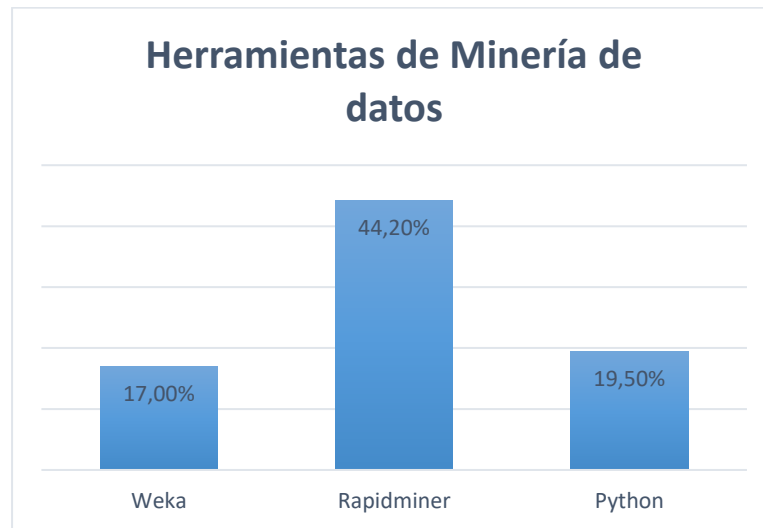


Figura 4. Herramientas para Minería de Datos obtenidas de KDnuggets

#### **4.6.1. Weka**

Weka es una plataforma de software utilizada para el aprendizaje supervisado y no supervisado, presenta una serie de algoritmos, cada uno con sus respectivos atributos y características. Además, proporciona un conjunto de librerías java para la extracción de conocimientos desde bases de datos [27].

#### **4.6.2. RapidMiner**

RapidMiner es una aplicación informática que sirve para el análisis y Minería de datos. Permite el desarrollo de procesos de análisis mediante el encadenamiento de operadores a través de un entorno gráfico [28].

#### **4.6.3. Spyder - Python**

Spyder es un IDE multiplataforma de código abierto para la ciencia de datos, integra las bibliotecas esenciales como NumPy, SciPy, Matplotlib e IPython, entre otras, además de eso, se puede ampliar con complementos.

#### 4.6.4. Pandas

Es una herramienta de análisis y manipulación de datos de código abierto, siendo rápida, potente, flexible y fácil de usar, construida sobre el lenguaje de programación Python [29], utilizada frecuentemente para la ciencia de datos.

#### 4.7. Metodología de Minería de Datos

Para llevar a cabo el desarrollo de un proyecto dentro de cualquiera que sea el ámbito, ya sea de ingeniería o cualquiera que sea su enfoque, es fundamental contar con una metodología de desarrollo, esta contará con un conjunto de actividades que generan documentación; dentro del proyecto el uso de la metodología no asegura la eficacia del proyecto, pero si permitirá administrar de forma eficiente el desarrollo del mismo, tomando en cuenta todos los requisitos y recursos destinados para este [30].

Si bien es cierto, existe una serie de metodologías que se utilizan para desarrollar Minería de Datos, entre ellas esta CRISP-DM, esta es una de las más utilizadas según KDnuggets <sup>2</sup>, el portal líder en IA, Analytics, Big data, Data mining, Data science, y Machine Learning, en donde realizaron muestran los resultados de una encuesta realizada en el 2007 y 2014, tal y como se muestra en la Figura 5.

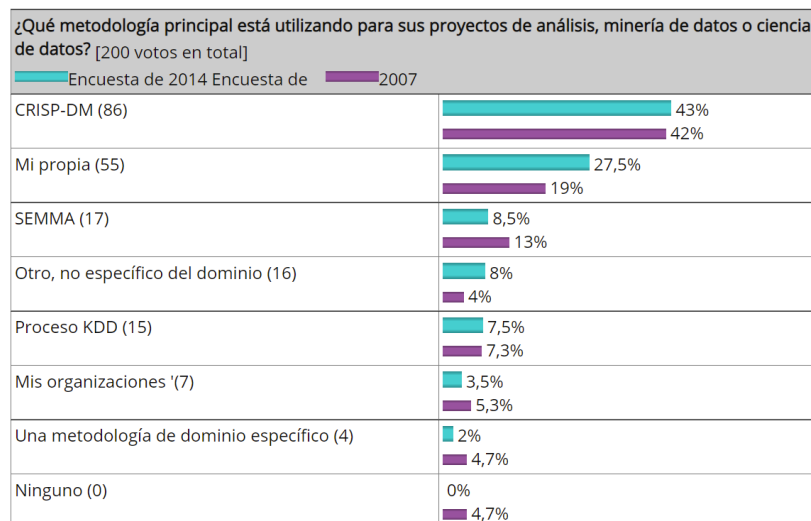


Figura 5. Resultados de CRISP-DM en encuesta realizada por KDnuggets (2014)

<sup>2</sup> <https://www.kdnuggets.com>

#### 4.7.1. CRISP-DM

Es una metodología de referencia para el desarrollo de proyectos de minería de datos más utilizada dentro de ambientes académicos e industriales [31]. CRISP-DM incluye un modelado y una guía, compuesta por seis fases, dentro de algunas de estas se puede regresar a una fase anterior para volver a revisarla, la Figura 6 muestra la secuencia de las fases de dicha metodología.



Figura 6. Fases de la Metodología CRISP-DM

Así mismo, la Figura 7 muestra a detalle cada una de las actividades necesarias para el cumplimiento de cada una de las fases de la Metodología CRISP-DM, así como la responsabilidad de cada fase:

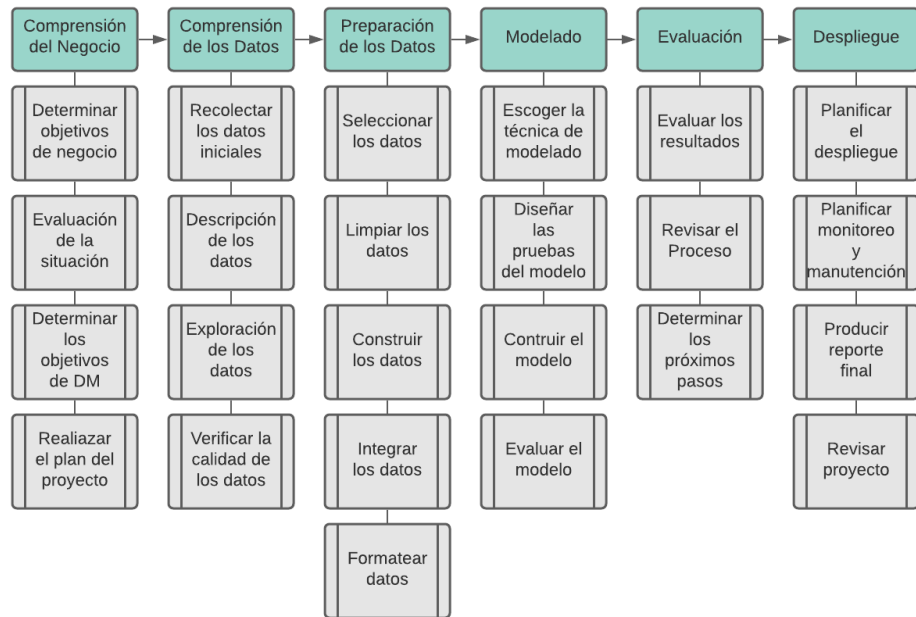


Figura 7. Desglose de las fases de la Metodología CRISP-DM

1. **Comprensión del Negocio:** Esta fase se centra en comprender cada uno de los objetivos y los requerimientos necesarios para la elaboración y ejecución de las actividades desde una perspectiva del negocio, con el propósito general de encontrar una solución que permita cumplir con cada uno de ellos de manera satisfactoria.
2. **Comprensión de los datos:** esta fase comprende la recolección inicial de los datos con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis.
3. **Preparación de los datos:** esta fase comprende en la preparación de los datos con los cuales se puedan trabajar con las técnicas de minería de datos seleccionada, para ello fue necesaria la estructura de datos de la fase anterior, de la misma manera se realiza una limpieza de los datos para poder generar el modelo de minería de datos, tratando de obtener una data sin valores nulos o vacíos.

4. **Modelado:** en esta fase se seleccionan las técnicas de minería de datos apropiadas para el TT, dentro de esto los parámetros son calculados a los valores más óptimos. Casi siempre existen varias técnicas para el mismo tipo o contexto del problema. Algunas de ellas requieren la forma de los datos de manera específica y ante esto se puede volver a la fase anterior.
  
5. **Evaluación:** dentro de esta fase se evalúa el modelo generado en la sección anterior. Es necesario revisar el proceso, considerando los resultados obtenidos, de tal forma que este lo más cercano posible al cumplimiento de los objetivos del negocio.
  
6. **Despliegue:** Una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, estos modelos se los integra dentro de los procesos de toma de decisiones de la organización [32], [33].

## 5. MATERIALES Y MÉTODOS

De acuerdo con el Reglamento de Régimen Académico que rige a las Instituciones de Educación Superior de Ecuador REC-SO-08-No- III-2019, en el artículo 32, numeral 3, se estipula que, para la aprobación de la unidad de integración curricular de deberá desarrollar un trabajo de integración curricular denominado anteriormente Trabajo de Titulación (TT), el cual se basará en procesos de investigación e intervención [34]. Todo TT deberá consistir en una propuesta innovadora que contenga como mínimo una investigación exploratoria y diagnóstica, además, de acuerdo con el artículo 40 del mismo reglamento, la investigación a nivel de grado es de carácter exploratorio y descriptivo [35]. La investigación exploratoria en el TT y a través de una RSL, se logró identificar los patrones que más destacaron dentro de los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas dentro de la Carrera de Ingeniería en Sistema/Computación. Con la obtención de la información arrojada mediante la minería de datos en dichos registros, se pudo realizar la elaboración de un plan de acción de acuerdo a las necesidades que se presentaron en los resultados, con ello se pretende ayudar a las autoridades correspondientes la toma de decisiones que impliquen mitigar la deserción de los estudiantes en Carreras de Ingeniería, así como también apoyando al proceso de enseñanza aprendizaje dentro de la Institución.

Así mismo, la investigación descriptiva brindó las respuestas necesarias para cada una de las preguntas implícitas en cada objetivo mediante su desarrollo y cumplimiento; y se encuentra sustentada en el apartado de Resultados.

Esta sección expone los materiales y métodos utilizados durante el desarrollo del TT. En el apartado 5.1 de detalla el contexto donde se realizó el TT, a continuación, en la sección 5.2 se muestra todo el proceso para cumplir con cada uno de los objetivos cada una con sus respectivas tareas: la sección 5.3 se presentan los recursos utilizados, y finalmente en la sección 5.4 de indican los participantes que intervinieron durante el TT.

## 5.1. Contexto

El presente Trabajo de Titulación se desarrolló dentro de la Universidad Nacional de Loja<sup>3</sup>, en la Facultad de la Energía, las Industrias y los Recursos Naturales No Renovables, en la Carrera de Ingeniería en Sistema/Computación.

## 5.2. Proceso

El proceso para alcanzar con el objetivo general del TT se detallan a continuación, mencionando a cada uno de los objetivos con sus respectivas actividades:

1. Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería.
  - Establecer métricas y lineamientos para la selección de trabajos relacionados con el tema de estudio.
  - Obtener y analizar los trabajos relacionados en base a métricas y lineamientos antes mencionados.
  - Elaborar Documentación.
2. Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.
  - Seleccionar la herramienta de minería de datos a utilizar para el desarrollo del trabajo.
  - Determinar el algoritmo más óptimo para la minería de datos.
  - Analizar e interpretar los resultados obtenidos.
3. Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.
  - Definir los objetivos del plan de acción.
  - Construir el plan de acción de acuerdo a los objetivos planteados.

---

<sup>3</sup> <https://unl.edu.ec/>



### **5.3. Recursos**

Para dar respuesta a la pregunta de investigación se hizo uso de recursos científicos como técnicos, éticos, académicos y metodológicos, los cuales se muestran a continuación:

#### **5.3.1. Recursos Científicos**

- Metodología para la RSL: para el desarrollo de una RSL objetiva se utilizó la metodología de Bárbara Kitchenham expuesta en [36], dando como resultado se obtuvo la revisión presentada en el Anexo 1.
- Estudios de casos: se lo utilizó para abordar los casos similares dentro de la RSL, ya que según [35] son estudios de campo u observaciones, con ello se logró comprender y sustentar de manera teórica el TT.
- Gestión de Proyectos PMBOOK: guía metodológica que proporcionó los pasos para la gestión de proyecto; el inicio y planificación se utilizó dentro del Anteproyecto, y ejecución, control y cierre se especifican en las secciones posteriores en el desarrollo del TT.
- Método científico: según [37], [38], se emplea con el fin de incrementar el conocimiento en ciencias, permitiendo mantener una familiaridad con el objeto de estudio mediante un análisis directo y el cual a través de una secuencia de etapas permitirá dar contestación a la pregunta de investigación dentro del TT. Las 4 etapas utilizadas fueron:

Observación: permitió conocer y analizar el problema y partir de ahí, empezar con el desarrollo del proceso de investigación. (Ver Anexo 21. Anteproyecto de Titulación)

Formulación de Hipótesis: esto permitió hacer una idea en base a la relación de ciertas variables presentes en el problema de investigación. El planteamiento de la hipótesis parte de la pregunta de investigación del problema (ver Anexo 21. Anteproyecto de Titulación). Para ello con el cumplimiento de los objetivos específicos se procede a dar a aceptar o rechazar a la hipótesis.

Experimentación: se realizó con el propósito de confirmar o rechazar la hipótesis planteada en esta investigación. El proceso de comprobación se aplicó dentro de cada uno de los objetivos específicos (ver Anexo 21. Anteproyecto de Titulación), los mismos cuentan con su respectiva documentación en la sección de resultados (ver Resultados)

Conclusiones: una vez que culminado el proyecto y verificada la hipótesis, se procedió a plantear las conclusiones y recomendaciones para trabajos a futuro (Ver Conclusiones, Recomendaciones y Trabajos Futuros).

### **5.3.2. Recursos Técnicos**

#### **Python**

Software que se utilizó para realizar el análisis estadístico de cada uno de las variables y conjunto de datos en general, además de la aplicación de los scripts para la determinación de patrones.

#### **Weka**

Esta plataforma sirvió para aplicar los diferentes algoritmos de minería de datos para poder determinar los patrones presentes en los registros académicos de los estudiantes.

#### **RapidMiner**

Software de entorno gráfico de fácil interacción con el usuario, donde se aplicaron los diferentes algoritmos de minería de datos y encontrar la información predominantes del conjunto de datos de los estudiantes.

#### **Google Colab**

Se lo utilizó para realizar los Script de Python para obtener las reglas de asociación y clústeres.

### **5.3.3. Recursos Éticos**

Consentimiento informado: se lo aplicó para contar la autorización de las personas entrevistadas para el desarrollo del plan de acción (Ver **Anexo 4. Ingeniería en Sistemas – Primer ciclo** hasta el **Anexo 17. Ingeniería en Computación – Cuarto ciclo**).

#### **5.4. Participantes**

El proceso de investigación fue realizado por Jhulissa Villamagua como Investigador, conjuntamente con el asesoramiento del Dr. Luis Chamba Eras (tutor del TT), el mismo que orientó y supervisó los aspectos técnicos y académicos durante todo el proceso de desarrollo del TT. El Ing. Edison Coronel, quién brindó asesoría académica acerca de las variables dentro del conjunto de datos original. Así mismo, se contó con la participación de la Dra. Marcia Criollo y de los Ingenieros Andrés Navas y Wilman Chamba, quienes aportaron al TT con la revisión, asesoramiento y validación de la propuesta del Plan de acción planteado por el Investigador.

## **6. RESULTADOS**

### **6.1. Fase 1: Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería.**

En esta primera fase se examinó el estado de la cuestión sobre los patrones predominantes que se presenta en registros académicos de los estudiantes de Ingeniería con asignaturas relacionadas con las matemáticas, el cual para su correcto desarrollo se basó en una Revisión Sistemática de Literatura (RSL), siguiendo a [36] y al enfoque presentado por Juan D. Velásquez en [39]–[41], así mismo se apoyó de herramientas para RSL descritas [42] tomando en consideración a la más óptima la herramienta denominada “Parsifal”<sup>4</sup>, ésta presenta sus servicios en línea diseñada con el propósito de ayudar a investigadores a realizar RSL. Tanto la documentación como la herramienta tecnológica se organizaron de manera correcta con cada una de las actividades planteadas dentro de esta primera fase, las mismas que analizaremos a detalle en los apartados a continuación y se pueden consultar en el **Anexo 1. Revisión sistemática de la literatura.**

#### **6.1.1. Establecer métricas y lineamientos para la selección de trabajos relacionados con el tema de estudio.**

Con el fin de llevar un desarrollo adecuado de la RSL se plantearon directrices (métricas y lineamientos) las cuales encaminan a la misma, estas directrices se manifiestan como preguntas de investigación, las cuales fueron contestadas al culminar todo el proceso que conlleva una RSL; se plantea 5 preguntas de investigación, donde ¿Cuáles son los factores más comunes presentes en la deserción estudiantil?”, el investigador podrá determinar cuáles son los factores más comunes que se presentan dentro de las instituciones que hacen que los estudiantes deserten o abandonen sus estudios; la segunda pregunta “¿Las asignaturas relacionadas con las matemáticas influyen en la deserción académica?” buscar dar a conocer al investigador sobre la relación que tiene la matemática con el desenvolvimiento del estudiante, de la misma manera busca dar a conocer si las matemáticas representan algún

---

<sup>4</sup> <https://parsif.al/>

obstáculo para que los estudiantes logren culminar sus estudios. La tercera pregunta, ¿Qué técnicas de Minería de Datos se utilizan para determinar factores que influyen en la deserción universitaria? Permite dar a conocer las técnicas más utilizadas que se implementan para el tratamiento de los datos y identificar cual es la más apropiada para este tipo de problemas. La cuarta pregunta, ¿Cuáles son las herramientas que se utilizan para el proceso de minería de datos?, encaminada a determinar cuál es la herramienta más utilizada para llevar a cabo la minería de datos. Y, finalmente ¿Cuáles son las principales metodologías para el desarrollo de minería de datos?, enfocándose en una metodología adecuada para el proceso de desarrollo de minería de datos y del contexto de la investigación.

Por otra parte, se establecen métricas para la selección de los trabajos, las cuales guiaron a la resolución de las preguntas de investigación, estas métricas establecen que la literatura seleccionada debe estar dentro de los últimos 5 años de publicación (2015-2020)<sup>5</sup>, donde los documentos serán únicamente obtenidos de fuentes bibliográficas confiables como son: Scopus, IEEE, ACM, Scielo, y para completar con la literatura gris dentro del TT se ha hecho uso de Google Académico.

#### **6.1.2. Obtener y analizar los trabajos relacionados en base a métricas y lineamientos antes mencionados.**

La obtención de bibliografía que apoya al cumplimiento de esta primera fase se desarrolló en base a las métricas y lineamientos establecidos en la **Sección 6.1.1**, de la misma manera se elaboraron palabras claves como: Minería de datos, deserción, abandono, matemáticas, dropout, student desertion, abandonment, mathematics, maths, Data Mining, influence, factors, engineering students, engineering, higher education, university, tools, education, weka, RapidMiner, Rstudio, Wolfram, Python, Knime, herramientas, metodología, methodology y con las palabras que anteceden se logró elaborar las cadenas de búsqueda. Finalmente, se identificó que la herramienta “Parsifal” al ingresar las palabras claves, se

---

<sup>5</sup> Fecha asignada en la que ejecutó la búsqueda

genera una cadena de búsqueda híbrida, por ello, se procedió a mejorarla y organizarla de tal manera que se adapte cada buscador seleccionado y de esta manera la búsqueda sea más exacta; estas cadenas de búsqueda se encuentran en la Tabla IX. del **Anexo 1. Revisión Sistemática de Literatura.**

Una vez aplicada la cadena de búsqueda acorde a cada buscador y utilizar los respectivos criterios de inclusión y exclusión, se obtuvo un total de 67 trabajos relacionados que responden a las preguntas de investigación. Estos se encuentran en las Tablas X-XII dentro del **Anexo 1. Revisión Sistemática de Literatura.**

Todo el proceso para la obtención de trabajos relacionados se sintetiza en el diagrama de Flujo expuesto en la Figura 8:

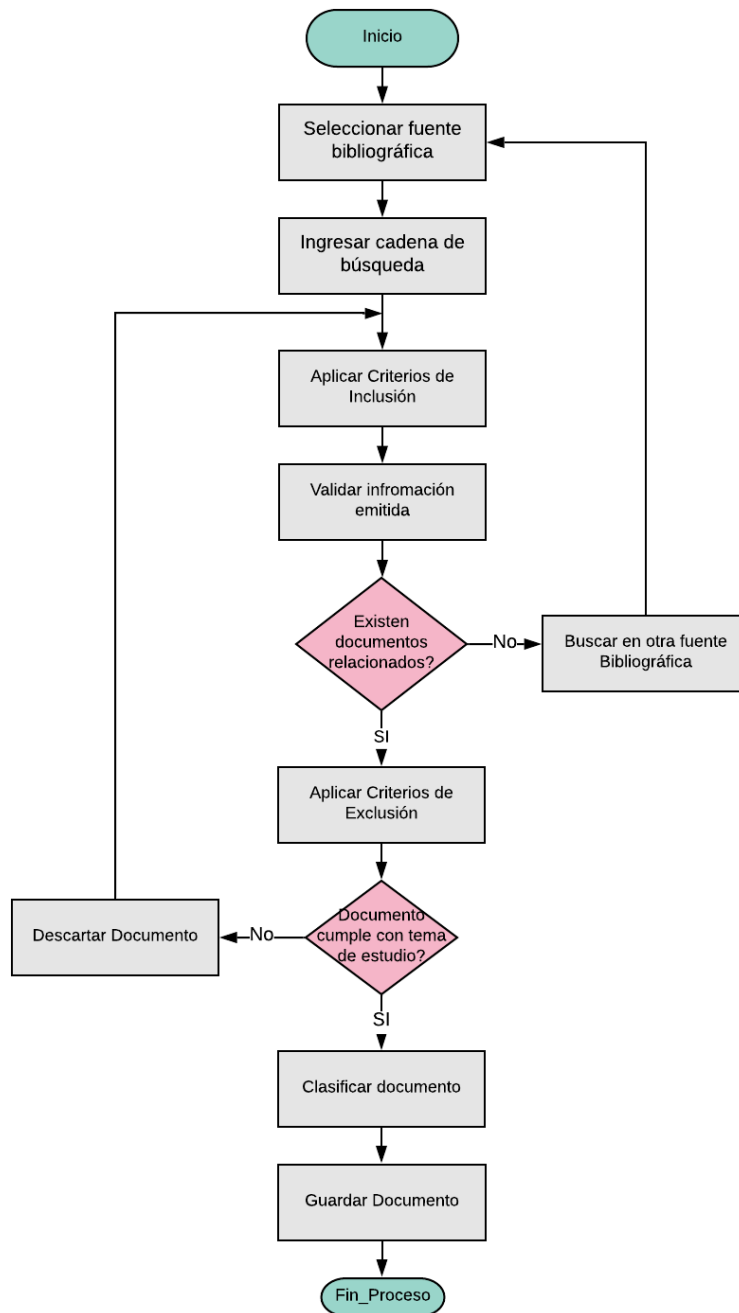


Figura 8. Diagrama de flujo para la obtención de documentos relacionados

Analizados cada uno de los trabajos relacionados se demuestra que, muchas de las investigaciones utilizan los registros académicos de la institución en estudio para tratar dicha información y encontrar soluciones que permitan determinar las causas o factores que influyen a muchos de los estudiantes a dejar o abandonar sus estudios, esto con el afán de mitigar este problema que asecha a muchas Universidades a nivel mundial.

Dentro de la misma, se puede determinar que el proceso de enseñanza dentro de Carreras de Ingeniería se muestra más robusto en comparación a otras Carreras, puesto que la presencia de las matemáticas dentro de estas es indispensable. De igual forma, se observa cierto auge en lo que corresponde “encontrar soluciones” dentro de Carreras de Ingeniería haciendo hincapié en asignaturas que se relacionen con las matemáticas, considerando que, es el área donde según investigaciones a nivel mundial se presenta un alto índice de deserción, enfatizando el bajo número de mujeres dentro de las mismas.

Por otro lado, dentro de los trabajos encontrados por parte de la RSL, se logró identificar que la técnica de Minería de Datos más predominante es de Aprendizaje Supervisado o Predictivo, siendo el algoritmo de decision tree en más implementado, sin embargo, este no muestra aporte a la investigación puesto que, se necesitan técnicas no supervisadas o descriptivas para determinar los patrones que predominan dentro del conjunto de datos por lo que, a través de [43][24], [26], [43], [44] se obtiene a las técnicas de Minería de datos descriptivas que se va a utilizar para tratar este tipo de datos, como son las Reglas de Asociación y Clústering. Tomando en consideración lo antes mencionado se muestra la necesidad de encontrar patrones presentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, y de esta manera identificar a los factores que indiquen en el alumno para desertar o abandonar sus estudios y con ello, plantear alternativas o estrategias con base a los resultados obtenidos que sirvan de apoyo a las autoridades correspondientes para la toma de decisiones y con ello tratar de mitigar el problema que se presenta en la Institución.

### **6.1.3. Elaborar Documentación.**

La documentación correspondiente a la primera fase se ve reflejada en la RSL presente en el **Anexo 1. Revisión sistemática de la literatura**, el cual detalla todas las actividades ejecutadas dentro de esta sección.



## **6.2. Fase 2: Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.**

En esta segunda fase se examinó el uso de ciertas herramientas con las que se puede obtener información relevante almacenada en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación. Adicionalmente, para su cumplimiento fue necesario conocer y analizar cada una de las variables o atributos pertenecientes al conjunto de datos con el que se va a desarrollar el TT, para ello se realizó un Análisis Exploratorio de los Datos, el cual proporcionó un conocimiento a nivel específico de todo el dataset, para posteriormente proceder a aplicar los algoritmos y determinar los patrones existentes dentro del mismo, esto se expone en el **Anexo 2. Análisis Exploratorio de los datos.**

### **6.2.1. Seleccionar la herramienta de minería de datos a utilizar para el desarrollo del trabajo.**

En esta actividad, con base a los resultados obtenidos por la RLS, existen muchos documentos que hacen referencia al uso de herramientas para la búsqueda de patrones. Dentro de los trabajos relacionados, la Tabla X hace referencia a Weka, RapidMiner y Python como las herramientas más utilizadas para determinar los patrones predominantes en el conjunto de datos, con base a las herramientas obtenidas, se ha logrado determinar a las herramientas a utilizarse para aplicar las técnicas de Minería de Datos Educativa, aclarando que existe una gran cantidad de herramientas que pueden servir para aplicar esta técnica, sin embargo, para el tema en estudio, estas técnicas y herramientas son las más implementadas para la determinación de patrones. La Figura 9 muestra algunas herramientas consideradas importantes para el presente Trabajo de Titulación, las mismas que se pueden visualizar a detalle en el **Anexo 1. Revisión Sistemática de Literatura.**

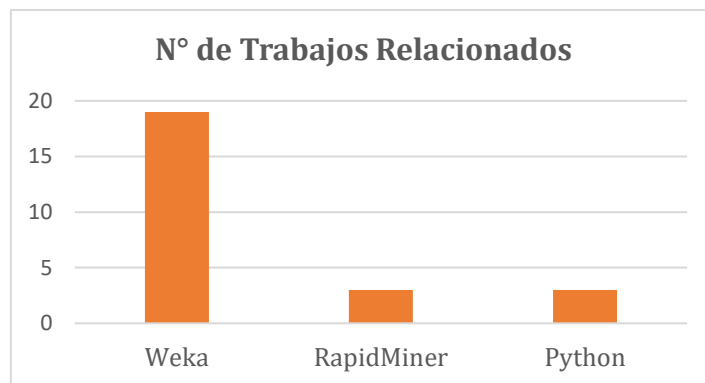


Figura 9. Herramientas a utilizarse para la Minería de Datos Educativa.

En la Figura 9 se muestra a la herramienta Weka como una de las más implementadas con 18 de 24 trabajos relacionados, RapidMiner con 3 de 24 trabajos obtenidos y, finalmente Python con 3 de 24 documentos. Con cada de una de estas herramientas se ha hecho la aplicación de los algoritmos, puesto que permitirán validar la calidad de los resultados obtenidos, y con base a su amplia variedad de funciones obtener un producto más cercano a la realidad, por lo que estas se convierten en las herramientas seleccionadas para el desarrollo del presente trabajo.

### 6.2.2. Determinar el algoritmo más óptimo para la minería de datos.

Como en la actividad anterior, también se recurre a los resultados obtenidos en la RSL, donde se aplicaron las interrogantes necesarias para el desarrollo del presente TT. En virtud de esto, se ha determinado que los trabajos obtenidos hacen referencia a algoritmos de clasificación, los cuales para la investigación no muestran la información que se requiere para la elaboración del plan de acción.

En relación a lo que antecede, se ha tomado de [23], [26], [43]–[46] las técnicas descriptivas a utilizarse para determinar la información relevante dentro del conjunto de datos como son las reglas de asociación y Clústering (agrupamiento) aplicadas al descubrimiento de información en los registros académicos de los estudiantes.

### 6.2.3. Analizar e interpretar los resultados obtenidos.

Una vez obtenido, las herramientas y los algoritmos a implementar para el tema de investigación, se procedió a ejecutar cada uno de las técnicas de minería de datos Educativa (técnicas descriptivas), como son Reglas

de Asociación y Clústering, producto de ellos se elaboran los scripts utilizados a lo largo del desarrollo del TT, mismos que se encuentran almacenados en el repositorio de GitHub<sup>6</sup>, para su replicabilidad y constancia. Así mismo, se aplicó una metodología que permita llevar una gestión del proyecto a desarrollar de manera ordenada y que, permita cubrir todos los aspectos que se requieren para el propósito del tema en estudio.

En virtud de lo antes mencionado, se ha hecho uso de la metodología CRISP-DM, la cual al ejecutar cada una de sus fases con sus respectivas actividades, se obtiene un reporte expuesto en el **Anexo 3. Reporte de Minería**, en el que se detalla todo el proceso de minería de datos desde el preprocesamiento de los datos hasta el resultado de los mismos.

Al mismo tiempo, ejecutados los algoritmos se obtiene la información que predomina dentro de los registros académicos de los estudiantes, vinculados a factores sociales, demográficos y académicos, en relación a cada ciclo y Carrera (Sistemas y Computación). Por tal motivo, se ha tomado como ejemplo al modelo obtenido por el quinto ciclo de la Carrera de Ingeniería en Sistemas, considerando que estos modelos como se ha mencionado con anterioridad, son en base a los ciclos de la carrera de Ingeniería en Sistemas y Computación, donde, existen 10 modelos de Sistemas y 4 de Computación, esto en relación a la fecha actual donde únicamente de tienen estudiantes hasta este ciclo. Los modelos restantes se encuentran a partir del **Anexo 4. Ingeniería en Sistemas – Primer ciclo** hasta el **Anexo 17. Ingeniería en Computación – Cuarto ciclo**.

La Figura 10 indica las características a procesar en la herramienta Weka, y con las mismas se procede a seleccionar el algoritmo de acuerdo al procesamiento.

---

<sup>6</sup> [https://github.com/Jhulissa/data\\_mining\\_for\\_student\\_dropouts](https://github.com/Jhulissa/data_mining_for_student_dropouts)

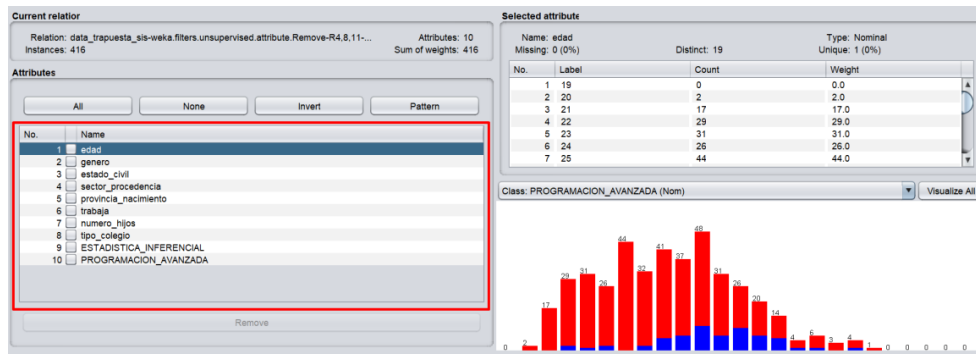


Figura 10. Selección de variables para la elaboración del modelo

Considerando al algoritmo A priori, se obtienen las siguientes reglas, mismas que se han obtenido aplicando los respectivos filtros, con nivel de confianza mayor a 0.9 se ha obtenido premisas mayores al 96% de confiabilidad y lift mayor a 1. Con ello se obtiene un total de 18 reglas (Ver Figura 11).

REGLAS	CONF.	LIFT.	LEV.	CONV.
1. genero=0 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 124 ==> PROGRAMACION_AVANZADA=1 122	conf:(0.98)>	lift:(1.14)	lev:(0.04)	conv:(5.66)
2. genero=0 sector_procedencia=0 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 111 ==> PROGRAMACION_AVANZADA=1 109	conf:(0.98)>	lift:(1.14)	lev:(0.03)	conv:(5.07)
3. genero=0 trabaja=0 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 102 ==> PROGRAMACION_AVANZADA=1 100	conf:(0.98)>	lift:(1.14)	lev:(0.03)	conv:(4.66)
4. genero=0 sector_procedencia=0 trabaja=0 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 90 ==> PROGRAMACION_AVANZADA=1 88	conf:(0.98)>	lift:(1.13)	lev:(0.02)	conv:(4.11)
5. genero=0 provincia_nacimiento=11 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 82 ==> PROGRAMACION_AVANZADA=1 80	conf:(0.98)>	lift:(1.13)	lev:(0.02)	conv:(3.75)
6. genero=0 sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 72 ==> PROGRAMACION_AVANZADA=1 70	conf:(0.97)>	lift:(1.13)	lev:(0.02)	conv:(3.29)
7. ESTADISTICA_INFERENCIAL=0 66 ==> trabaja=0 64	conf:(0.97)>	lift:(1.08)	lev:(0.01)	conv:(2.33)
8. genero=0 provincia_nacimiento=11 trabaja=0 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 66 ==> PROGRAMACION_AVANZADA=1 64	conf:(0.97)>	lift:(1.12)	lev:(0.02)	conv:(3.01)
9. genero=1 tipo_colegio=1 80 ==> trabaja=0 77	conf:(0.96)>	lift:(1.08)	lev:(0.01)	conv:(2.12)
10. genero=0 ESTADISTICA_INFERENCIAL=1 248 ==> PROGRAMACION_AVANZADA=1 238	conf:(0.96)>	lift:(1.11)	lev:(0.06)	conv:(3.09)
11. genero=1 sector_procedencia=0 tipo_colegio=1 74 ==> trabaja=0 71	conf:(0.96)>	lift:(1.07)	lev:(0.01)	conv:(1.96)
12. tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 191 ==> PROGRAMACION_AVANZADA=1 183	conf:(0.96)>	lift:(1.11)	lev:(0.04)	conv:(2.91)
13. provincia_nacimiento=11 trabaja=0 tipo_colegio=2 PROGRAMACION_AVANZADA=1 114 ==> sector_procedencia=0 109	conf:(0.96)>	lift:(1.04)	lev:(0.01)	conv:(1.46)
14. provincia_nacimiento=11 trabaja=0 tipo_colegio=2 136 ==> sector_procedencia=0 130	conf:(0.96)>	lift:(1.04)	lev:(0.01)	conv:(1.49)
15. genero=0 sector_procedencia=0 ESTADISTICA_INFERENCIAL=1 226 ==> PROGRAMACION_AVANZADA=1 216	conf:(0.96)>	lift:(1.11)	lev:(0.05)	conv:(2.82)
16. provincia_nacimiento=11 trabaja=0 tipo_colegio=2 ESTADISTICA_INFERENCIAL=1 112 ==> sector_procedencia=0 107	conf:(0.96)>	lift:(1.03)	lev:(0.01)	conv:(1.44)
17. genero=1 tipo_colegio=1 ESTADISTICA_INFERENCIAL=1 67 ==> trabaja=0 64	conf:(0.96)>	lift:(1.07)	lev:(0.01)	conv:(1.77)
18. trabaja=0 tipo_colegio=2 156 ==> sector_procedencia=0 149	conf:(0.96)>	lift:(1.03)	lev:(0.01)	conv:(1.5)

Figura 11. Reglas obtenidas aplicando filtros (Conf – Lift)

Algunas de las reglas son:

1. Si un estudiante es de género masculino, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
2. Si un estudiante es de género masculino, procedente del sector Urbano, no ejerce actividad laboral, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que también apruebe la materia de Programación Avanzada.
3. Si un estudiante es de género femenino, culmina sus estudios secundarios en una Institución Fiscal y aprueba la asignatura de

Estadística Inferencial es muy probable que sea debido a que el estudiante no ejerce actividad laboral.

### Resultados obtenidos por Algoritmo A priori – RapidMiner

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las únicas asignaturas relacionadas a las matemáticas se mantiene, a estas se suman variables como el tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 23 a 28 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 12.

```
[PROGRAMACION AVANZADA, edad, tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.780)
[edad, tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.792)
[tipo_colegio, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.814)
[PROGRAMACION AVANZADA, edad] --> [ESTADISTICA INFERENCIAL] (confidence: 0.826)
[genero] --> [PROGRAMACION AVANZADA] (confidence: 0.829)
[genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.829)
[PROGRAMACION AVANZADA, tipo_colegio, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.833)
[PROGRAMACION AVANZADA, edad, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.836)
[tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.837)
[tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.846)
[tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.856)
[ESTADISTICA INFERENCIAL, tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.857)
[ESTADISTICA INFERENCIAL, edad, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.868)
[sector_procedencia] --> [ESTADISTICA INFERENCIAL] (confidence: 0.875)
[sector_procedencia] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.875)
[ESTADISTICA INFERENCIAL, edad, tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.885)
[PROGRAMACION AVANZADA, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.892)
[ESTADISTICA INFERENCIAL, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.892)
[sector_procedencia] --> [PROGRAMACION AVANZADA] (confidence: 0.906)
[PROGRAMACION AVANZADA, tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.907)
[ESTADISTICA INFERENCIAL, edad] --> [PROGRAMACION AVANZADA] (confidence: 0.911)
[PROGRAMACION AVANZADA] --> [ESTADISTICA INFERENCIAL] (confidence: 0.916)
[ESTADISTICA INFERENCIAL, tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.918)
[trabaja] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.932)
[ESTADISTICA INFERENCIAL] --> [PROGRAMACION AVANZADA] (confidence: 0.940)
[provincia_nacimiento] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.953)
[trabaja] --> [PROGRAMACION AVANZADA] (confidence: 0.955)
[trabaja] --> [ESTADISTICA INFERENCIAL] (confidence: 0.955)
```

Figura 12. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio, como lo muestra la Figura 13, se ha demostrado que para aprobar a las materias

relacionadas con las matemáticas, como son Estadística Inferencial y Programación, deben estar vinculadas con las variables tales como: no ejercer ninguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que las asignaturas sean aprobadas el estudiante debe dedicarse únicamente a las actividades académicas que le corresponden, de estado civil soltero, procedentes de sector Urbano y finalmente, la correlación que existe entre las materias de este ciclo.

```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardado...
+ Código + Texto
RelationRecord(items=frozenset({'EI_A', 'PA_A'}), support=0.7908653846153846, ordered_statistics=[OrderedStatistic(it
RelationRecord(items=frozenset({'EI_A', 'Soltero'}), support=0.7884615384615384, ordered_statistics=[OrderedStatistic
RelationRecord(items=frozenset({'Soltero', 'PA_A'}), support=0.7932692307692307, ordered_statistics=[OrderedStatistic
RelationRecord(items=frozenset({'EI_A', 'Soltero', 'PA_A'}), support=0.7475961538461539, ordered_statistics=[OrderedS
RelationRecord(items=frozenset({'EI_A', 'PA_A', 'Urbano'}), support=0.7235576923076923, ordered_statistics=[OrderedSt
RelationRecord(items=frozenset({'EI_A', 'Soltero', 'Urbano'}), support=0.7235576923076923, ordered_statistics=[Ordere
RelationRecord(items=frozenset({'NTrabaja', 'Soltero', 'PA_A'}), support=0.7019230769230769, ordered_statistics=[Orde
RelationRecord(items=frozenset({'Soltero', 'PA_A', 'Urbano'}), support=0.7259615384615384, ordered_statistics=[Ordere

```

Figura 13. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 14.

Clustered Instances		
0	236	( 57%)
1	83	( 20%)
2	33	( 8%)
3	27	( 6%)
4	37	( 9%)

Figura 14. Clústeres con sus respectivas instancias

Dentro de estos, el clúster con más características agrupadas es el 0, debido a que cuenta con el mayor número de observaciones, es decir, 236 observaciones, mientras que, el 3 es el clúster cuenta con 27 observaciones siendo este el menos representativo. La Figura 15, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Full Data (416.0)	Cluster#				
		0 (236.0)	1 (83.0)	2 (33.0)	3 (27.0)	4 (37.0)
edad	29	25	28	23	22	29
genero	0	0	1	0	0	1
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	1	0	0
tipo_colegio	1	2	1	1	2	1
ESTADISTICA_INFERENCIAL	1	1	1	1	1	1
PROGRAMACION_AVANZADA	1	1	1	1	1	1

Figura 15. Clústeres obtenidos por Weka.

**Clúster\_0 (57%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_1 (20%):** Estudiantes que tengan una edad de 28 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_2 (8%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, cuenta con una actividad laboral y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

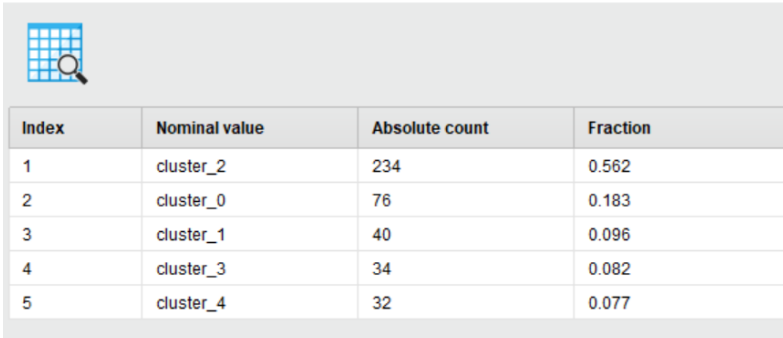
**Clúster\_3 (6%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán

todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_4 (9%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figura 16 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústeres.



The screenshot shows a window titled "Nominal values" with a close button (X) in the top right corner. Inside the window, there is a search icon (magnifying glass over a grid) and a table with the following data:

Index	Nominal value	Absolute count	Fraction
1	cluster_2	234	0.562
2	cluster_0	76	0.183
3	cluster_1	40	0.096
4	cluster_3	34	0.082
5	cluster_4	32	0.077

Figura 16. Clúster con total de instancias

Por otra parte, en la Figura 17, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 416 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ESTADISTIC...	PROGRAMA...
1	1	cluster_4	-0.327	1.542	3.460	-0.125	-0.344	0.912	0.434	0.398
2	2	cluster_2	1.574	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
3	3	cluster_3	-1.413	-0.647	-0.288	2.500	-0.344	-0.861	0.434	0.398
4	4	cluster_2	-0.599	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
5	5	cluster_2	-0.055	-0.647	-0.288	-1.438	-0.344	-0.861	0.434	0.398
6	6	cluster_2	0.216	-0.647	-0.288	-1.438	-0.344	-0.861	0.434	0.398
7	7	cluster_2	-0.599	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
8	8	cluster_2	-0.055	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
9	9	cluster_2	-1.142	-0.647	-0.288	-0.125	-0.344	0.912	0.434	0.398
10	10	cluster_0	0.488	1.542	-0.288	-0.125	-0.344	0.912	-2.300	-2.507
11	11	cluster_2	-1.413	-0.647	-0.288	-0.125	-0.344	0.912	0.434	0.398
12	12	cluster_2	-1.685	1.542	-0.288	-0.125	-0.344	-0.861	0.434	0.398
13	13	cluster_3	-0.599	-0.647	-0.288	1.844	-0.344	0.912	0.434	0.398

Figura 17. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 18 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 2 con más instancias agrupadas.

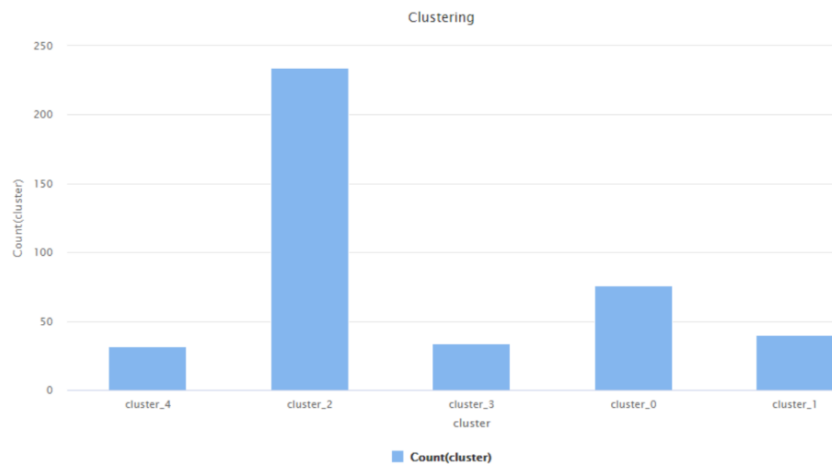


Figura 18. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes (Ver Figura 19).

### k-Means - Centroid Table

Cluster	edad	ESTADISTICA L...	genero	PROGRAMACI...	provincia_nac...	sector_proce...	tipo_colegio	trabaja
Cluster 0	26.634	0.976	0.146	0.976	11.927	0	1.439	1
Cluster 1	27.094	0.875	0.219	0.906	11.625	1.000	1.469	0.094
Cluster 2	27.359	0.978	1	0.891	11.772	0	1.370	-0
Cluster 3	26.173	1	0	0.953	11.277	-0	1.565	-0
Cluster 4	30.700	0.017	0.300	0.433	10.617	0	1.450	-0

Figura 19. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 26 años, es de género masculino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscomisional y que cuenta con una actividad laboral extra a partir de sus actividades académicas.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 27 años, es de género masculino, perteneciente a la provincia de Loja y al sector Rural, proviene de un colegio Fiscal y que no cuenta con una actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 27 años, es de género femenino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscal y que no cuenta con una actividad laboral.

**Clúster 3:** Estudiante aprueba únicamente la asignatura de Geometría plana y reprueba Cálculo Diferencial y Fundamentos Informáticos, todas pertenecientes al primer ciclo, si tiene una edad de 26 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 4:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 30 años,

es de género masculino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscal y no cuenta con una actividad laboral.

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica Figura 20, se han obtenido los clústers con las variables Género y Tipo de colegio, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.

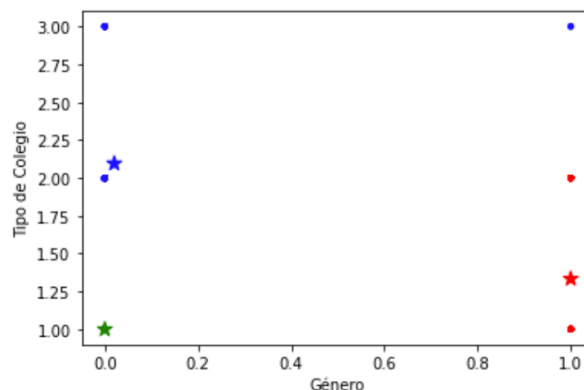


Figura 20. Relación Género – Tipo de Colegio

En la Figura 21, se muestra la relación entre la edad y la asignatura Metodología de la Programación, en donde los estudiantes aprueban a la asignatura si su edad es de 24, 28 y 33 años. Cabe recalcar que

estudiantes con edad de 33 años, pueden probabilísticamente perder a dicha asignatura.

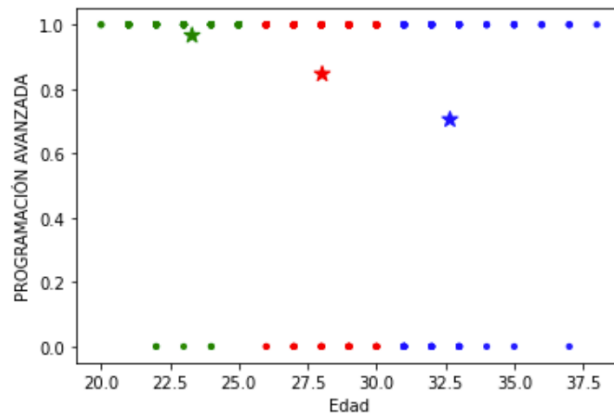


Figura 21. Relación Edad – Programación Avanzada

Por otra parte, dentro de la Figura 22 se muestra la relación entre el Género y Programación Avanzada, donde se observa que un grupo mínimo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género con probabilidad de reprobar la materia de Programación Avanzada.

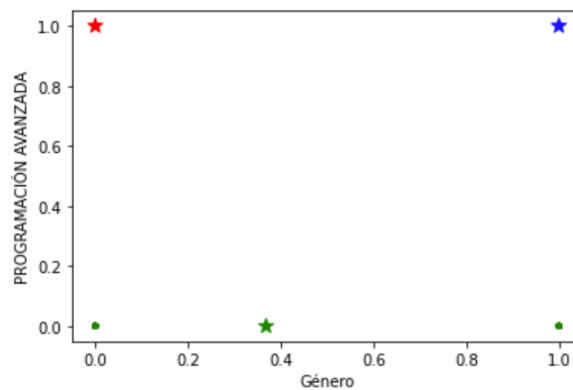


Figura 22. Relación Género - Programación Avanzada

Finalmente, la Figura 23 hace referencia a la relación que se da entre materias, en este caso Estadística Inferencial y Programación Avanzada, en un clúster (verde) se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde muy probable se apruebe a Estadística Inferencial pero reprobará Programación avanzada, mientras que aprobará

Programación Avanzada pero no tendrá ninguna posibilidad de aprobar Estadística inferencial.

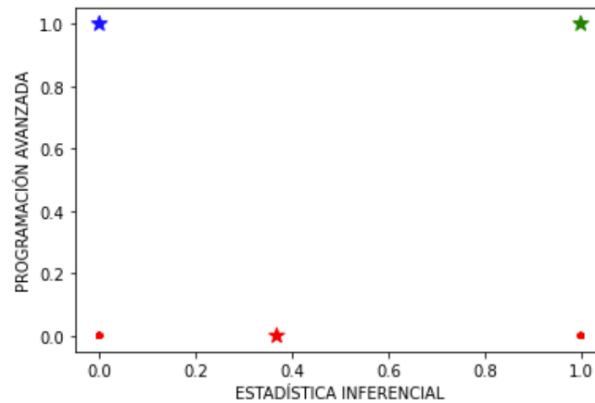


Figura 23. Relación Estadística Inferencial - Programación Avanzada

Adicionalmente, se ha obtenido a las características que predominan en el quinto ciclo de acuerdo a la confianza tal y como se muestra en la Figura 24, donde se identifica que, estudiantes que no ejercen ninguna actividad laboral, es una de las características más representativas en cuanto a variables sociales; otra de las características con valores elevados son para aquellos estudiantes que provienen del sector Urbano, así como también, hace referencia a instituciones Fiscales y Fiscomisionales donde los estudiantes culminaron sus estudios secundarios, esto como parte de las variables demográficas, para este ciclo se identifica la presencia del género femenino como antecedente, sin embargo en este ciclo sigue predominado el género masculino en conjunto con las variables antes mencionadas, finalmente para las características más representativas dentro de las variables académicas, se muestra a la materia de Estadística Inferencial como la más predominante en ambos estados, aprobada y reprobada, mientras que, la aprobación de la asignatura de Programación Avanzada se da como consecuente, esto quiere decir que necesita de otros parámetros para ser aprobada con éxito.

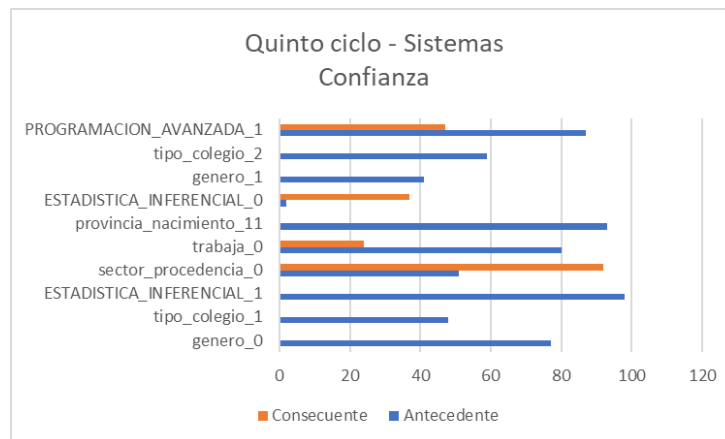


Figura 24. Confianza - Quinto ciclo Sistemas

Cada uno de los modelos obtenidos hacen referencia a una serie de variables sociales, demográficas y académicas, mismas que fueron encontradas a partir de la aplicación de las técnicas de minería de datos; a nivel específico, se hace referencia a características trabajar, provincia de nacimiento, edad, género, Institución donde culminaron sus estudios secundarios, entre otros atributos destacados dentro de la minería de datos. Todos los resultados correlacionados, se encuentran a detalle en el **Anexo 3. Reporte de Minería de Datos**, y dentro de cada uno de la minería aplicada a cada ciclo.

### 6.3.Fase 3: Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.

Obtenidos los resultados de la fase anterior, donde se identifica que, los estudiantes desertan de la Carrera por contar con una actividad laboral, algunos estudiantes cuentan con un compromiso, es decir, son casados o se encuentran en unión libre lo que los compromete a buscar un sustento económico, lo cual da resultado a que la mayoría de estudiantes aprueban las asignaturas si no cuentan con una actividad laboral y son de estado civil soltero; pertenecer al sector Rural es otro de los factores por los que los estudiantes desertan ya que, las bases de conocimiento que se adquieren son inferiores a los que se obtienen en el sector Urbano, asimismo, la institución donde los estudiantes culminan sus estudios secundarios forma parte de las causas de abandono, ya que, las bases académicas necesarias para que el estudiante continúe con su formación

académica universitaria son muy bajas. Por otra parte, se ha identificado que la mayor parte de la comunidad universitaria (estudiantes) son provenientes específicamente de las provincias de Loja, El Oro y Zamora Chinchipe. Finalmente, dentro de la información obtenida se refleja que, aunque el problema de deserción se presenta en el género masculino es mucho más evidente en el género femenino. Las causas son diversas, pero esta es la información que se ha obtenido aplicada la minería de datos educativa a los registros académicos de los estudiantes de Ingeniería en Sistemas y Computación, con lo cual se ha desarrollado una propuesta de plan de acción para mitigar la deserción universitaria que sirva de apoyo a las autoridades correspondientes para la toma de decisiones.

#### **6.3.1. Definir los objetivos del plan de acción.**

Con la finalidad de brindar apoyo a la problemática detectada en la Universidad Nacional de Loja sobre los altos índices de deserción en Carreras de Ingeniería, se propone un plan de acción que permitirá diseñar las acciones y/o estrategias necesarias para lograr que la Institución mantenga la permanencia de los estudiantes dentro de las Carreras. El plan de acción servirá como base para el cambio y su propósito será apoyar y orientar a las autoridades correspondientes a la toma de decisiones y a su vez propiciará al cumplimiento y finalidad por el cual fue creada la misión y visión de la Universidad.

La propuesta se basa en la planificación, coordinación, capacitación y recursos, de tal manera que cada uno de estos elementos sirvan de orientación para todo el personal que interviene en el plan de acción, mismos que deben mostrar un comportamiento interdisciplinario donde la toma de decisiones para minorar los índices de deserción en la UNL se tome de forma consensuada, así como también muestren un desempeño competente y que estén encaminados al cumplimiento de los objetivos de la Institución.

## **Objetivo del Plan de Acción**

Definir estrategias que permitan mitigar los índices de deserción universitaria y mejorar la permanencia de los estudiantes en la Carrera de Sistemas/Computación.

## **Objetivos Específicos**

1. Concientizar a estudiantes, docentes y áreas allegadas a trabajar en conjunto para mitigar los altos números de deserción que se da en la UNL.
2. Brindar un enfoque general sobre la misión, los objetivos y a los aprendizajes a adquirir en la Carrera que se escogerá.
3. Reforzar las bases académicas con las que ingresan los estudiantes a la Universidad.
4. Elaborar talleres y/o comunidades que permitan a los estudiantes reforzar los conocimientos obtenidos en clase.
5. Coordinar con las áreas correspondientes para brindar alternativas de apoyo económico a los estudiantes.
6. Incentivar al género femenino a formar parte de Carreras STEAM.
7. Mejorar el material o las técnicas didácticas de enseñanza utilizadas para impartir clases.
8. Solicitar apoyo psicológico que motive a los estudiantes a alcanzar sus propósitos.
9. Impulsar la participación de los estudiantes en los diferentes congresos o concursos que se dicten en la UNL
10. Mantener un registro donde se pueda visualizar el avance de cada una de las actividades.

### **6.3.2. Construir el plan de acción de acuerdo a los objetivos planteados.**

Para la construcción del plan de acción se basó en la estructura de [47],[48], la misma que se adaptó a las necesidades del TT. Una vez establecidos los objetivos del plan de acción, se consideran los siguientes aspectos para el plan de acción.



**Fase:** Hace referencia a la estructura del plan de acción, la secuencia que debe seguir para alcanzar el objetivo, en este caso las fases son: sensibilizar, desarrollo de actividades y seguimiento y control.

**Problemática:** indica el conjunto de causas o complicaciones que originan la deserción académica.

**Objetivos Específicos:** son aquellas metas a perseguir y con base a esto se plantean las estrategias que apoyarán al cumplimiento del objetivo principal.

**Metas:** es lo que se espera conseguir una vez aplicadas las estrategias descritas en cada objetivo.

**Estrategias:** son las acciones planificadas que se utilizarán para la toma de decisiones y de esta manera mejorar las vulnerabilidades que se presentan en la Carrera de Ingeniería en Sistemas/Computación.

**Recursos:** son los elementos que ayudarán a ejecutar las estrategias planteadas, estos son Humanos y técnicos.

**Tiempo:** es el periodo durante el cual se va a ejecutar las acciones del plan de acción. Este depende de una previa planificación.

**Responsable:** persona(s) encargada de la planificación, ejecución y control de cada una de las actividades descritas en el plan de acción.

A continuación, se muestra la propuesta del plan de acción para mitigar la deserción académica universitaria dirigido a la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja.

Tabla I. Plan de acción

<p align="center"><b>“UNIVERSIDAD NACIONAL DE LOJA”</b>  <b>PROPUESTA DE PLAN DE ACCIÓN PARA MITIGAR LA DESERCIÓN UNIVERSITARIA</b></p>							
<b>OBJETIVO:</b>		“Definir estrategias que permitan mitigar los índices de deserción universitaria y mejorar la permanencia de los estudiantes en la Carrera de Sistemas/Computación”					
<b>FASE</b>	<b>PROBLEMÁTICA</b>	<b>OBJETIVOS ESPECIFICOS</b>	<b>METAS</b>	<b>ESTRATEGIAS</b>	<b>RECURSOS</b>	<b>TIEMPO</b>	<b>RESP.</b>
1. SENSIBILIZACIÓN	Falta de conocimiento e información al personal involucrado sobre las medidas necesarias para minorar la deserción universitaria.	1. Concientizar a estudiantes, docentes y áreas allegadas a trabajar en conjunto para mitigar los altos números de deserción que se da en la UNL.	Incentivar a todo el personal involucrado a tomar las medidas de acción necesarias para mitigar la deserción en la UNL.	<ul style="list-style-type: none"> <li>➤ Dar a conocer a nivel general el propósito y la importancia de crear un plan de acción que sirva de apoyo en la toma de decisiones a todo el personal involucrado.</li> </ul>	<p><b>Humano:</b> personal involucrado (docentes, estudiantes, bienestar estudiantil, entre otros que se consideren necesarios para lograr el objetivo principal.</p> <p><b>Material:</b> Plan de acción.</p>	Jornada de 45 min a una hora.	Gestor de la Carrera de Sistemas/ Computación
2. DESARROLLO DE ACTIVIDADES	Falta de interés en los estudiantes que ingresan a la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja	1. Brindar un enfoque general sobre la misión, los objetivos y a los aprendizajes a adquirir en la Carrera que se escogerá.	Determinar el interés estudiantil por la Carrera que va a escoger.	<p>Contrarrestar el bajo interés de los estudiantes en la Carrera mediante:</p> <ul style="list-style-type: none"> <li>➤ Coordinar docentes en conjunto con estudiantes las visitas a los colegios para brindar</li> </ul>	<p><b>Humanos:</b> Docentes, estudiantes, Unidad de Telecomunicación e Información (UTI).</p>	<p>Durante el último curso de colegio.</p> <p>Casa abierta, durante media mañana</p>	Universidad Nacional de Loja.

				<p>información sobre la carrera.</p> <ul style="list-style-type: none"> <li>➤ Implementar casas abiertas donde se proporcione la debida información y conocimiento general sobre la carrera.</li> <li>➤ Crear Webinars dirigido a estudiantes de secundaria en donde se muestren las diferentes Carreras que oferta la UNL con un enfoque de las mismas, así como también las áreas de trabajo donde se pueden desempeñar.</li> </ul>	<p><b>Materiales:</b> Computadoras, proyectores, Flyers publicitarios, trípticos, Software, plazas para casa abierta.</p>		
<p>Los conocimientos adquiridos por las Instituciones Fiscales y en caso particulares Fiscomisionales no se adaptan a los requerimientos solicitados por la Universidad.</p>	<p>2. Reforzar las bases académicas con las que ingresan los estudiantes a la Universidad.</p>	<p>Conseguir que los estudiantes fortalezcan todos los conocimientos básicos como por ejemplo matemática, física, lógica, y</p>	<ul style="list-style-type: none"> <li>➤ A través de la prueba de diagnóstico disponer de manera obligatoria a docentes, para despejar las dudas que se presentan en los alumnos y, a estudiantes cuyo</li> </ul>	<p><b>Humanos:</b> Docentes, estudiantes.</p> <p><b>Materiales:</b> Computadoras, proyectores, cuaderno, lápices, etc.</p>	<p>Dos semanas después de los resultados de la prueba de diagnóstico</p>	<p>Docentes, estudiantes.</p>	

			<p>habilidades como resolución de problemas, perseverancia, orden, entre otros que se necesitan en la Carrera a seguir y con ello evitar el abandono de la misma.</p>	<p>resultado de la misma es bajo, asistan de manera obligatoria a las tutorías impartidas por los maestros.</p> <ul style="list-style-type: none"> <li>➤ Trabajar más en grupos colaborativos entre estudiantes, para compartir conocimientos y generar un ambiente de confianza y compañerismo.</li> </ul>			
<p>Durante el avance del ciclo académico se muestran “debilidades” en ciertas materias que pueden ser perjudiciales para la aprobación del ciclo del estudiante.</p>	<p>3. Elaborar talleres y/o comunidades que permitan a los estudiantes reforzar los conocimientos obtenidos en clase.</p>	<p>Mantener un buen porcentaje de continuidad en los estudiantes que avanzan al siguiente semestre.</p>	<ul style="list-style-type: none"> <li>➤ Conformar grupos de estudio y trabajo académico estudiantiles.</li> <li>➤ Realizar actividades que evalúen los contenidos teóricos y los mismos sean aplicados en prácticas relacionados a la vida profesional y su aplicación con otras áreas.</li> <li>➤ Crear comunidades de acorde a las necesidades de los</li> </ul>	<p><b>Humano:</b> Docentes, estudiantes, ponentes externos</p> <p><b>Materiales:</b> Salsa de reuniones (aulas) computador personal, proyectores, libros, copias, etc.</p>	<p>Jornadas de dos o tres días a la semana con un promedio de horas similar (dos o tres h/día).</p>	<p>Presidente de la Carrera en conjunto con Gestor de la carrera.</p>	

				estudiantes que les permita reforzar los conocimientos adquiridos en clases, los cuales pueden ser impartidos por estudiantes de ciclos superior, docentes u personal externo que desee colaborar.			
	Unos de los factores para la deserción académica son las variables económicas, en las cuales aquellos estudiantes que cuentan con un trabajo, descuidan sus estudios y esto conlleva a la pérdida o abandono de la carrera.	4. Coordinar con las áreas correspondientes para brindar alternativas de apoyo económico a los estudiantes.	Evitar que los estudiantes abandonen sus estudios por cumplir con una actividad laboral.	<ul style="list-style-type: none"> <li>➤ Validar información emitida por los estudiantes sobre la ficha socioeconómica a través de la intervención de un trabajador social, y brindar opciones de permanencia del estudiante mediante becas, auxilios o subsidios económicos.</li> <li>➤ Trabajar en conjunto con la UTI, para brindar asesoría y soporte técnico en actividades básicas a la ciudadanía en</li> </ul>	<p><b>Humano:</b> Estudiantes, área intermediaria (bienestar estudiantil), Institución beneficiaria, Unidad de Telecomunicación e Información (UTI)</p> <p>.</p> <p><b>Materiales:</b> Área para soporte , kit de herramientas, computadoras, documentos, copias, etc.</p>	Gestión de visitas y trámite correspondiente Externo.	<p>Gestor de la Carrera de Sistemas/ Computación.</p> <p>Trabajador social (visitas)</p>

				<p>general con el propósito de generar ingresos económicos que servirán de apoyo para el estudiante.</p> <ul style="list-style-type: none"> <li>➤ Charlas socioeconómicas que incentiven a los estudiantes a permanecer en la carrera.</li> </ul>			
	<p>El porcentaje de estudiantes de mujeres en la Carrera es bajo, puesto que, desde instancias familiares entre otros causales califican a estas como Carreras para hombres, y durante el proceso de aprendizaje ciertas materias les resultan complicadas y difíciles de entender, lo mismo conlleva al abandono o cambio de la Carrera por otras.</p>	<p>5. Incentivar al género femenino a formar parte de Carreras STEAM.</p>	<p>Lograr que el género femenino rompa estereotipos y forme parte de Carreras STEAM.</p>	<ul style="list-style-type: none"> <li>➤ Brindar apoyo a estudiantes en asignaturas relacionadas a la resolución de problemas a través de un computador en las que presenten inconvenientes.</li> <li>➤ Crear programas que, de forma creativa y divertida fomenten el interés por la ciencia y las matemáticas entre los estudiantes.</li> <li>➤ Buscar referentes de mujeres informáticas y promocionarlas,</li> </ul>	<p><b>Humano:</b> Docentes, estudiantes, Organizaciones sin fines de lucro que apoyen la iniciativa.</p> <p><b>Materiales:</b> computador, proyector, libros, etc.</p>	<p>Dos días por semana y de ser el caso coordinar con Institución para horario).</p> <p>Jornada de 1 o 2 horas</p>	<p>Presidente de la Carrera.</p>

				<p>dando a conocer la igualdad de género en Carreras STEAM.</p> <ul style="list-style-type: none"> <li>➤ Charlas que incentiven a las alumnas a mantenerse dentro de la Carrera y seguir formando parte del cambio.</li> <li>➤ Trabajar en conjunto con el género masculino en la ejecución de proyectos innovadores, valorando así el aporte femenino dentro de los mismos.</li> </ul>			
La falta de interacción entre el docente y el estudiante y la metodología empleada para llegar a los estudiantes son algunas de las causas por las que el alumno deserta la carrera.	6. Mejorar el material o las técnicas didácticas de enseñanza utilizadas para impartir clases.	Alejar el desinterés y conseguir la atención y entusiasmo de los estudiantes al recibir clases manteniendo la interacción maestro-	<ul style="list-style-type: none"> <li>➤ Implementar canales de comunicación confiable para mayor interacción con el estudiante.</li> <li>➤ Capacitar a los docentes tanto en metodología como en el uso de nuevas</li> </ul>	<p><b>Humano:</b> docentes, estudiantes.</p> <p><b>Materiales:</b> Computador, proyecto, ,software intuitivo y amigable.</p>	Durante cada clase impartida.	Docente de asignatura	

			estudiante y viceversa.	<p>herramientas tecnológicas para impartir clases.</p> <ul style="list-style-type: none"> <li>➤ Aplicar nuevas metodologías de enseñanza que permitan al estudiante obtener un nivel competente elevado.</li> <li>➤ Utilizar software intuitivo para realizar actividades en clases, prácticas, tareas, etc.</li> </ul>			
Un indicador primordial es el estado de ánimo en los estudiantes, puesto que hace referencia a lo que está pasando por la vida del estudiante, esto influye a que los estudiantes no presten la atención ni el entusiasmo que se requiere en sus estudios.	7. Solicitar apoyo psicológico que motive a los estudiantes a alcanzar sus propósitos.	Lograr la permanencia de los estudiantes sin que los problemas sean un impedimento para lograr su objetivo.	<ul style="list-style-type: none"> <li>➤ Crear con el apoyo de Bienestar estudiantil temas generales a impartir, en lo que los estudiantes presentes vulnerabilidades.</li> <li>➤ Implementar a nivel de Carreras, una encuesta anónima con las preguntas necesarias que permitan conocer un resultado concreto y poder</li> </ul>	<p><b>Humano:</b> Estudiantes, Bienestar Estudiantil. Unidad de Telecomunicación e Información (UTI)</p> <p><b>Materiales:</b> software, redes sociales.</p>	Una encuesta semestral o anual.	El servicio prestado debe estar habilitado todo el tiempo.	Bienestar estudiantil.



				<p>brindar una solución, no sin antes dar a conocer la importancia de la misma y el apoyo de la Institución.</p> <ul style="list-style-type: none"> <li>➤ Ofrecer servicios gratuitos de apoyo psicológico a estudiantes.</li> </ul>			
<p>La participación de los estudiantes en los actos que propicia la Universidad es baja, no existe mucha interacción con otras Carreras, lo puede resultar un impedimento para la creación de nuevas propuestas.</p>	<p>8. Impulsar la participación de los estudiantes en los diferentes congresos o concursos que se dicten en la UNL</p>	<p>Incitar a los estudiantes a presentar proyectos y a su vez integrar a Carreras STEAM con otras Carreras.</p>	<ul style="list-style-type: none"> <li>➤ Potenciar e incentivar a los docentes para que se presenten proyectos en los que pueda intervenir los estudiantes, reforzar conocimiento y aprendizaje.</li> <li>➤ Utilizar los recursos con los que cuenta la Universidad para la elaboración de los proyectos.</li> <li>➤ Promover y ejecutar las propuestas generadas por los estudiantes con el apoyo de la UTI o de entidades en convenio como parte de las</li> </ul>	<p><b>Humano:</b> Estudiantes, Bienestar Estudiantil. Unidad de Telecomunicación e Información (UTI)</p> <p><b>Materiales:</b> software</p>	<p>Depende de la propuesta presentada</p>	<p>Docentes, estudiantes, líder de proyecto.</p>	

				prácticas preprofesionales.			
3. SEGUIMIENTO Y CONTROL		Mantener un registro donde se pueda visualizar el avance de cada una de las actividades		➤ Elaborar un cronograma con criterios de aceptación o entregables para la ejecución de cada una de las actividades.	<b>Humano:</b> Estudiantes, Docentes, Gestor de la Carrera, Bienestar Estudiantil. Unidad de Telecomunicación e Información (UTI),  <b>Materiales:</b> software	Participantes de cada actividad. Gestor de la carrera  <b>Materiales:</b> documentos, cámara fotográfica.	Reuniones de 1 o 2 horas trimestrales o por cada final de semestre.  Visitas al lugar donde se efectúa la actividad (si amerita)

## **7. DISCUSIÓN**

El desarrollo del presente TT se basa en tres objetivos específicos, encaminados al cumplimiento del objetivo general, a continuación, se detalla la discusión por cada uno de los objetivos. La sección 7.1 expone los resultados contrastándolos con la literatura relacionada con el objeto de estudio; la sección 7.2 presenta la valoración científica, técnica, económica y ambiental del TT y finalmente, la sección 7.3 las limitaciones que se presentaron en el desarrollo del presente TT.

### **7.1. Desarrollo de la propuesta alternativa**

#### **7.1.1. Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería**

Para el cumplimiento del primer objetivo, se realizó una RSL (Ver Anexo 1), esta fase permitió familiarizarse con los conceptos necesarios en cuanto a minería de datos, algoritmos, metodología y herramienta a utilizarse acorde a las necesidades de la investigación. Dentro de los resultados obtenidos por la RSL se obtuvieron un conjunto de trabajos enfocados a determinar las causas de abandono académico universitario, así como también a buscar una solución que sirva de apoyo para conseguir la permanencia de los alumnos en las Universidades, para ello se hace uso de los datos académicos almacenados en el sistema con el que cuenta cada institución, mientras que, en otros casos para tratar este inconveniente, recurren a la obtención de la información necesaria, como son entrevistas, encuestas, entre otras, esto con el fin de aplicar las diferentes técnicas de la Inteligencia Artificial y obtener resultados que persigan el mismo objetivo enfocado a que el número de desertores académicos minore considerablemente. De la misma manera, algunos de los trabajos se encuentran realizados a nivel general, es decir, a nivel de Ingenierías [49], [50], mientras que en [51] se hace énfasis en algunas Carreras técnicas como son Arquitectura, Ingeniería Agrícola, entre otras. Asimismo, destacando a las matemáticas en las Ingenierías, no se ha logrado identificar un vínculo entre las mismas, en lugar de ello [52], [53] realizan estudios de manera individual, sin considerar que puede significar

un factor desertor muy importante que debe ser tratado. Por otra parte, dentro de los trabajos obtenidos ninguno hace referencia al uso específico de asignaturas que se relacionen a las matemáticas de acuerdo al código de la UNESCO como se propone en este caso, al implementar este tipo de limitaciones se puede determinar a fondo cuales son las asignaturas en las que los estudiantes presentan dificultades y que los orillan a abandonar su carrera.

Es importante conocer porque existen altos números de abandono dentro de estas Carreras, así como también conocer cuál es la relación con las matemáticas, esto permitirá conocer si es cuestión de metodología aplicada por el docente, gusto, temor, falta de comprensión por la misma, entre otros factores que influyen al abandono universitario.

Algunas de las limitaciones que se han presentado en el TT, se vinculan directamente con la falta de relación que existen entre las matemáticas y la Ingeniería considerando que estas van de la mano, de igual forma sucede con los trabajos que se vinculen con Ingeniería en Sistemas/Computación en particular, existen muy pocos estudios que se dediquen específicamente a tratar a las Carreras antes mencionadas.

#### **7.1.2. Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.**

Si bien es cierto existen varios trabajos que tratan de especificar algunas de las causas por las que existe la deserción universitaria, muchos de ellos describen a manera específica cada uno de los factores. Dentro de [3], [54]–[58] obtenidos por la RSL enfatizan a factores económicos, académicos, sociales como los principales causantes de abandono universitario, cada uno de ellos aplicando diferentes técnicas de Inteligencia Artificial, mismas que en su mayoría son técnicas de clasificación y dada que estas son técnicas predictivas no permiten determinar la información (patrones) predominante que se oculta dentro de los registros académicos.

En virtud de lo expuesto, a través de [23], [26], [43]–[46] se tomaron las técnicas descriptivas que se aplicaron para el descubrimiento de información en los registros académicos de los estudiantes del Ingeniería en Sistemas/Computación, de los cuales se identifica a edad, en donde se obtiene un promedio a nivel de cada ciclo, el género, siendo los hombres con el mayor número de observaciones así como también el sector Urbano, la etnia mestiza y no ejercer ninguna actividad laboral. Asimismo, las provincias más destacadas dentro del premisas y clúster obtenidos son la Provincia de Loja, El Oro y finalmente Zamora Chinchipe. Por último, el tipo de institución en la que los estudiantes culminan sus estudios secundarios, según [59] indica que los estudiantes con mejores resultados académicos son los que pertenecen a instituciones Particulares y algunas Fiscomisionales; en la Carrera de Ingeniería en Sistemas las Instituciones Fiscales y Fiscomisionales sobresalen en los resultados adquiridos, mientras que Computación los Colegios Fiscomisionales y Particulares son los que predominan. Sin embargo, hay ciertas características que son tratadas de manera particular, como por ejemplo la inclusión del género femenino dentro de Carreras de Ingeniería, así mismo, en relación al tipo de colegio están las bases que obtienen los estudiantes previo al ingreso universitario; Todos los resultados expuestos con anterioridad se encuentran expuestos en el **Anexo 3. Reporte de Minería**, dentro del cual redirecciona a cada uno de los resultados obtenidos por los diferentes ciclos (**Anexo 4. Ingeniería en Sistemas – Primer ciclo, Anexo 17. Ingeniería en Computación – Cuarto ciclo**) resultado de la aplicación de los algoritmos, estas guardan una similitud con los resultados obtenidos en la RSL. En relación a lo anteriormente expuesto, la información que prevalece en los trabajos relacionados obtenidos a nivel nacional e internacional es la misma que se obtiene en el desarrollo de este trabajo investigativo, puesto que la situación por la que cruza la UNL no se aleja del problema presente en otras instituciones.

La limitación más importante que se encontró durante el desarrollo de esta fase es la inconsistencia de los datos, esto tomó tiempo debido al estado

bruto en la que se encontraban los registros académicos de los estudiantes, es decir, este no se encontraba de la mejor manera y al momento de realizar la limpieza de los mismos fue necesario la implementación de técnicas predictivas que permitieron completar a los mismos, de igual forma, se utilizó el método de prueba y error para poder obtener un mejor resultado y así mantener la veracidad de los datos. Por otra parte, existen pocos trabajos que trabajen con técnicas descriptivas, ya que la mayoría están enfocadas a técnicas de predicción. Finalmente, el número de características obtenidas por cada ciclo en particular los registros de la Carrera Computación son inferiores a 100 registros y, si bien es cierto no existe un número en particular como regla definido para efectuar la Minería de datos, es necesario contar con varias observaciones mismas que ayudarán a conocer la realidad de mejor manera.

### **7.1.3. Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.**

Para esta fase, en [56], [60], [61], mencionan una serie de resultados que guardan similitud con los resultados obtenidos en el presente TT, es decir factores sociales, demográficos y académicos. En base a estos, parte la elaboración de la propuesta del plan de acción que servirá de apoyo para mitigar la deserción académica en la UNL. Durante la RSL y la problemática identificada no se ha encontrado ningún trabajo en donde se proponga un plan de acción o de mejora que cuente con alternativas o estrategias que permitan disminuir el abandono de una Carrera universitaria. La propuesta del Plan de acción para el TT de titulación se basa en determinar las problemáticas principales que se presentan con base a los resultados obtenidos y, a partir de ellas, plantear un objetivo que contará con sus respectivas metas, estrategias, recursos, tiempo de ejecución y responsable del mismo. Asociados a los resultados se ha elaborado un conjunto de estrategias que ayuden a cumplir el objetivo planteado, sin embargo, al no existir referentes para el desarrollo de la propuesta del plan de acción este resultó ser una limitante para esta fase

debido a que no se tenían referencia de apoyo para elaborar las estrategias que se vinculen con el tema de estudio.

## **7.2. Valoración técnica, económica y ambiental**

Para la elaboración del presente Trabajo de Titulación fue necesario hacer uso de recursos técnicos, económicos y ambientales.

### **7.2.1. Valoración Técnica**

Los recursos técnicos aplicados para el TT se basan en la utilización de ciertas plataformas de software, IDE's las cuales sirvieron para aplicar las diferentes técnicas de minería de datos y obtener los modelos con los que se trabajó.

### **7.2.2. Valoración Económica**

Si bien es cierto durante el desarrollo del presente TT se utilizaron recursos que no implicaban gastos, sin embargo, por el contrario, también existieron recursos que implicaban gastos, los cuales se resumen en este apartado.

Tabla II. Recursos humanos

<b>Personal</b>	<b>Tiempo (Horas)</b>	<b>Precio/Hora (\$)</b>	<b>Valor total (\$)</b>
Investigador (alumno)	480	\$ 5,00	\$ 2,400
Director de TT	50	\$ 10,00	\$ 500,00
Docente	480	\$ 10,00	\$ 4,800
<b>Subtotal</b>			<b>\$ 7,700</b>

La Tabla II, muestra los recursos humanos que fueron necesarios para el desarrollo del presente TT, en este caso únicamente se contó con el investigador y el director de TT.

Con relación a los recursos materiales, a continuación, en la Tabla III se detallan todos los instrumentos que fueron necesarios para la elaboración de la memoria y del proyecto en general.

Tabla III. Recursos materiales

<b>Recursos</b>	<b>Cantidad</b>	<b>Valor Unitario (\$)</b>	<b>Valor total (\$)</b>
Internet	12 meses	\$ 30,00	\$ 360,00
Transporte	320 días	\$ 0.30	\$ 96,00
Suministros de Oficina	--	\$ 110	\$ 110,00
<b>Subtotal</b>			<b>\$ 566,00</b>

Seguidamente, la Tabla V, detalla los recursos técnicos y tecnológicos que se utilizaron para el desarrollo del TT.

Tabla IV. Recursos técnicos

<b>Recursos</b>	<b>Cantidad</b>	<b>Valor Unitario (\$)</b>	<b>Valor total (\$)</b>
<b>Hardware</b>			
Computador	1	\$ 1,200	\$ 1,200
Impresiones	3	\$ 50,00	\$ 150,00
Pendrive	1	\$ 20,00	\$ 20,00
<b>Software</b>			
Weka	1	\$ 0,00	\$ 0,00
RapidMiner	1	\$ 0,00	\$ 0,00
Python	1	\$ 0,00	\$ 0,00
<b>Subtotal</b>			<b>\$ 1,370</b>

Finalmente, se presenta a la Tabla V con el presupuesto aproximado necesario para el desarrollo de TT.

Tabla V. Costo aproximado del proyecto

<b>Recursos</b>	<b>Subtotal</b>
Recursos Humanos	\$ 7,700.00
Recursos Materiales	\$ 566,00
Recursos Tecnológicos	\$ 1,370.00
<b>SUBTOTAL</b>	<b>\$ 9,636.00</b>



Gastos imprevistos (5%)	\$ 481,80
<b>TOTAL</b>	<b>\$ 10, 117.80</b>

### 7.2.3. Valoración Económica

La valoración ambiental nace del uso de los diferentes recursos digitales presentes durante el TT, permitiendo así el ahorro de materiales educativos, como son impresiones, solicitudes, entre otras, que perjudican al medio ambiente.

### 7.3. Limitaciones

Bajo el desarrollo del TT, se han presentado aspectos que quedaron fuera de la cobertura del mismo, debido a que no formaron parte del alcance de los objetivos, pero es importante identificarlos ya que sirven de base para los trabajos futuros.

#### 7.3.1. Utilizar nuevas técnicas de Minería de Datos

Existen varios métodos (técnicas y algoritmos) que se pueden aplicar al proceso de Minería de Datos, considerando que la elección de los mismos depende del problema en estudio y de las variables que contenga el dataset, por tal motivo la elección del métodos van relacionados con el objetivo planteado para el tema de estudio. Bajo el desarrollo del TT, se han utilizado algoritmos para obtener información predominante dentro de los registros académicos, es decir, técnicas descriptivas, sin embargo, dada la inconsistencia de los mismos, es necesario aplicar o adaptar el conjunto de datos hacia técnicas predictivas, las mismas que brindarán un enfoque a futuro y a su vez, servirá como sustento a los patrones que se han determinado y de esta manera extrapolar al mismo.

#### 7.3.2. Aplicabilidad de Minería de Datos a todas las asignaturas

Bajo la ejecución del desarrollo del TT, se ha determinado que existe una correlación entre las materias relacionadas con las matemáticas, donde se refleja una dependencia para aprobar o reprobar a las mismas, sin embargo, al seleccionar únicamente las materias que se describen con anterioridad, esto se volvió una limitante ya que se descarta la probabilidad de encontrar la información o relación con todas las asignaturas y con los atributos sociales, demográficos y académicos, y de

esta manera comparar y contrastar con los resultados obtenidos. Así mismo, es necesario abordar más información de los ciclos superiores de la Carrera de computación, es decir, aplicar los modelos obtenidos a todos los registros académicos de los estudiantes de dicha carrera.

### **7.3.3. Población no contemplada**

Se dispone de tanta información que a veces es imposible organizarla con efectividad y obtener un modelo eficiente. Sin embargo, en ciertas ocasiones el porcentaje de información que se utiliza para aplicar las técnicas de minería de datos es muy baja y aunque esto no representa un impedimento, es necesario obtener un alto número de observaciones para que, al aplicar las técnicas de Minería de Datos el modelo obtenido sea más exacto. Durante el desarrollo del TT se ha trabajado únicamente con los ciclos disponibles de la Carrera de Computación, es decir, con los cuatro primeros ciclos, y por tal motivo, se ha convertido en una limitante ya que no se ha considerado todos los ciclos en la ejecución de la investigación, por tal motivo es necesario extrapolarlo a los nuevos ciclos así como también a los nuevos estuantes que forman parte de la Carrera de Ingeniería en Computación, para encontrar la información relevante dentro de los nuevos registros de los estudiantes.

## 8. CONCLUSIONES

Una vez culminado el TT, se obtiene las siguientes conclusiones:

- Las técnicas de Minería de Datos Educativa como son las Reglas de Asociación y Clúster, aplicadas en las herramientas Weka, RapidMiner y Python, permitieron identificar a factores tales como: no trabajar, pertenecer al sector Urbano, ser oriundo de la Provincia de Loja, de estado civil soltero y estudiantes de género masculino, como los más predominantes dentro de los registros académicos de los estudiantes de la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, relacionando las asignaturas al código de la UNESCO, identificando a estas como una de las causas que provoca el abandono de dichas Carreras ya que, en la mayoría de ciclos, la relación que guardan consigo mismas y con las características descritas con anterioridad, son muy dependientes.
- Los resultados de la RSL, hacen referencia a una serie de factores que intervienen en la deserción universitaria, estos están relacionados con sociales, demográficos y académicos, siendo este último con más observaciones. De la misma manera, algunos de los trabajos buscan obtener información con los registros académicos de los estudiantes, pero la mayoría parte con la recolección de información de encuestas aplicadas a estudiantes de Ingeniería. Con base a los 67 resultados analizados se logró obtener los conocimientos y bases necesarias para abordar el tema en estudio y dar contestación a las preguntas de investigación, conocer los factores más comunes para desertar la Carrera universitaria, algunas de las asignaturas específicas para tratar, las técnicas de minería de datos a utilizar, la metodología a seguir para el desarrollo del TT y la herramienta a aplicar para la obtención de resultados. Por otra parte, de los trabajos obtenidos por la RSL, ninguno hace referencia al uso específico de asignaturas que se relacionen a las matemáticas tomando como limitante el uso del código de la UNESCO; al implementar este tipo de condiciones se puede determinar a fondo cuales son las dificultades que se les presentan a los estudiantes para desertar su Carrera en relación con dichas asignaturas y, con base a los resultados obtenidos brindar un

aporte en la toma de decisiones para mitigar la deserción académica dentro de la UNL.

- Los patrones obtenidos al aplicar las diferentes técnicas de minería de datos con más observaciones son los correspondientes a variables académicas, sociales y demográficas, específicamente, se cuenta con estudiantes de género masculino, que no ejerzan ninguna actividad laboral, que cuenten con una edad media, es decir, jóvenes de entre 18 a 30 años, que pertenezcan al sector Urbano, a una de las provincias más representativas como son Loja, El Oro y Zamora Chinchipe, de estado civil soltero y que hayan culminado sus estudios secundarios en una Institución Fiscal o Fiscomisional; para la representación de las asignaturas, únicamente las relacionadas a las matemáticas, los resultados varían y dependen de cada ciclo, ya que cada uno representa un ambiente diferente y por ende diferentes asignaturas.
- El plan de acción propuesto para mitigar la deserción en la UNL, fue elaborado de acuerdo a los resultados obtenidos al aplicar las técnicas de minería de datos educativa, en el cual se menciona a tres fases, sensibilización, desarrollo de actividades y seguimiento y control con los siguientes cada una de estas cuenta con la problemática, los objetivos específicos, las metas a las que se espera llegar, las estrategias a implementar para cumplir el propósito del plan de acción, los recursos tanto humanos como técnicos, el periodo de tiempo y finalmente el responsable de realizar dicha acción o actividad, todos los aspectos necesarios para abordar a la problemática determinada con base a los factores predominantes determinados en el desarrollo del TT.

## 9. RECOMENDACIONES

- Con base a trabajos relacionados y determinadas las causas de la deserción universitaria basados en técnicas predictivas, es recomendable realizar un análisis de cada uno de los factores y las técnicas implementadas que han dado resultados positivos, extrapolando los resultados del TT con los obtenidos por las nuevas técnicas utilizadas.
- Con relación al conjunto de datos con el que se trabajó, es recomendable levantar nueva información que permita obtener características más específicas relacionadas con el abandono de la carrera, por ejemplo, afectación por ausencia de padres, familia disfuncional, problemas de salud, psicológicos, económicos, y/o afiliación a la carrera, ya sea de forma general o específica, es decir, por determinadas materias. Así mismo, se recomienda aplicar las técnicas de minería de datos encaminado a encontrar información única y relevante en todas las asignaturas que se presentan en cada ciclo académico en la que el estudiante presente más inconvenientes.
- De acuerdo con la información obtenida al aplicar las técnicas de Minería de Datos Educativa, específicamente en la Carrera de Computación, es recomendable aplicar dichas técnicas a la nueva población que no fue contemplada durante el desarrollo del TT, realizar una comparativa y adaptar o determinar diferentes estrategias que en base a los resultados obtenidos permitan mantener a los estudiantes dentro de la carrera.
- Es necesario ejecutar el plan de acción y que este no quede como una propuesta, más bien, dado los motivos por los que los estudiantes desertan la Carrera y considerando que este es solo el inicio para lograr el cambio, se debe trabajar en conjunto para mantener la permanencia del estudiante no solo dentro de Carreras de Ingeniería si no en cualquier otra, independiente si los resultados obtenidos no se relacionan con los trabajos a realizarse.

## TRABAJOS FUTUROS

- Con fines académicos es posible aplicar este trabajo investigativo en otras Carreras o su vez en las Facultades de la Universidad, para determinar cuáles son las causas de abandono que en las mismas predominan y presentar las estrategias adaptables a los resultados obtenidos.
- Con base al desarrollo del TT, es factible replicar el proceso realizado de Minería de Datos Educativa a través de técnicas de aprendizaje supervisado o predictivas y que, en apoyo de los resultados obtenidos los mismos que garantizando su reproductividad se encuentran almacenados en GitHub<sup>7</sup> se pueda manejar de mejor manera este fenómeno que se presenta en la UNL.
- Realizar una solución de Software que permita predecir con base a los resultados obtenidos el porcentaje de permanencia y/o abandono de los estudiantes que ingresan a las nuevas Carreras que presta la UNL.
- Recopilar nueva información o características que permita conocer si existen causas diversas a las que se presentan en el TT con base a encuestas o entrevistas y poder replicar el TT a las observaciones que se obtienen con los resultados obtenidos.

---

<sup>7</sup> [https://github.com/Jhulissa/data\\_mining\\_for\\_student\\_dropouts](https://github.com/Jhulissa/data_mining_for_student_dropouts)

## 10. BIBLIOGRAFIA

- [1] G. J. Zambrano, K. G. Rodríguez, and L. H. Guevara, “Análisis de la Deserción Estudiantil en las Universidades del Ecuador y América Latina.,” pp. 1–28, 2018.
- [2] F. Bravo, L. Illescas, S. Larriva, and M. Peña, “Causas de Deserción en el Ingreso a la Universidad,” *Rev. la Fac. Ciencias Químicas Univ. Cuenca*, vol. 18, pp. 48–59, 2017.
- [3] E. V. García, R. C. Rodríguez, S. C. Tumbaco, Y. P. Santana, R. L. Fernández, and L. B. Benítez, “Factores que influyen en la deserción de los estudiantes en la Universidad de Guayaquil,” *Rev. la Fac. Ciencias Médicas la Univ. Guayaquil*, vol. 0, no. 0, pp. 21–26, 2017.
- [4] L. del C. Guevara, “Estudio de las causas que generan la deserción académica de los estudiantes de la escuela de Psicología,” 2017.
- [5] N. S. Ortiz, X. F. Trujillo, and S. E. Tello, “Análisis de la deserción estudiantil en la carrera de ingeniería en teleinformática para proponer nueva metodología de enseñanza,” *CISCI 2018 - Decima Septima Conf. Iberoam. en Sist. Cibern. e Informatica, Decimo Quinto Simp. Iberoam. en Educ. Cibern. e Informatica, SIECI 2018 - Memorias*, vol. 2, pp. 177–182, 2018.
- [6] D. A. Becerra, “Estudio del Rendimiento Académico Aplicando Técnicas de Minería de Datos,” 2015.
- [7] A. Peña Fernández, “Los factores pedagógicos influyen en la deserción universitaria,” *INNOVA Res. J.*, vol. 4, no. 3, pp. 108–115, 2019, doi: 10.33890/innova.v4.n3.2019.996.
- [8] C. Calderón Mayorga, E. Colomo Magaña, and J. Ruíz Palmero, “Causas de la deserción escolar en Ingeniería, Electrónica y Computación del Centro Universitario de los Valles de la Universidad de Guadalajara (México).,” *Rev. Espac.*, vol. 41, no. 6, p. 14, 2020.
- [9] D. Mato, C. Soneira, and M. Muñoz, “Estudio de las Actitudes hacia las Matemáticas en Estudiantes Universitarios,” *Numer. Rev. Didact. las Mat.*, vol. 97, no. 1, pp. 7–20, 2018.
- [10] “Ineval presentó resultados de PISA-D | Instituto Nacional de Evaluación Educativa.”
- [11] “Resultados Ser Bachiller,” 2018.

- [12] G. M. López, A. C. Espinosa, V. M. López, and S. P. T. Cuenca, "Ser Bachiller 2018-2019," p. 19, 2019.
- [13] L. Antonio, J. Luis, and J. Iván, "La educación STEM y su potencial en la formación integral del talento universitario para la Industria 4.0 en Ecuador.," pp. 1–14, 2018.
- [14] H. E. Viale Tudela, "Una Aproximación Teórica a La Deserción Estudiantil Universitaria," *Rev. Digit. Investig. en Docencia Univ.*, no. 1, p. 59, 2014, doi: 10.19083/ridu.8.366.
- [15] Silva Alegría Tania Guadalupe and E. Aguirre, "Análisis de la incidencia del rendimiento académico en la deserción escolar de estudiantes de 3° A del turno matutino del Centro Escolar Público Nueva Vida en el distrito X del municipio de Ciudad Sandino del departamento de Managua durante el segundo sem.," vol. 23, no. 45, pp. 5–24, 2016.
- [16] J. C. Suárez, "Factores que generan miedo, apatía o desinterés frente al estudio de las Matemáticas," 2014.
- [17] P.M. González Urbaneja, "Historia de la matemática: Integración cultural de las matemáticas, génesis de los conceptos y orientación de su enseñanza," *Enseñanza las ciencias Rev. Investig. y Exp. didácticas*, vol. 9, no. 3, pp. 281–289, 1991.
- [18] R. D. Moreno Padilla, "La llegada de la inteligencia artificial a la educación," *Rev. Investig. en Tecnol. la Inf.*, vol. 7, no. 14, pp. 260–270, 2019, doi: 10.36825/riti.07.14.022.
- [19] C. B. Lenniet and P. G. O. Lidia, "Contribuciones de la Inteligencia Artificial a la Educación Superior," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013, doi: 10.1017/CBO9781107415324.004.
- [20] G. de la C. León Rodríguez and S. M. Viña Brito, "La inteligencia artificial en la educación superior. Oportunidades y amenazas.," *INNOVA Res. J.*, vol. 2, no. 8.1, pp. 412–422, 2017, doi: 10.33890/innova.v2.n8.1.2017.399.
- [21] A. Ballesteros, D. Sánchez-Guzmán, and R. García, "Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo.," *Latin-American J. Phys. Educ.*, vol. 7, no. 4, pp. 662–668, 2013.
- [22] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to



- 2005,” *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007, doi: 10.1016/j.eswa.2006.04.005.
- [23] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.
- [24] J. Molina and J. García, “Técnicas de Minería de Datos basadas en Aprendizaje Automático,” *Técnicas de Análisis de Datos*, pp. 96–266, 2008.
- [25] M. B. Beltrán, “Minería de Datos,” p. 67, doi: 10.1016/0032-0633(82)90071-X.
- [26] M. N. Moreno García and V. F. López Batista, “Uso de Técnicas no Supervisadas en la Construcción de Modelos de Clasificación en Ingeniería del Software,” *Dep. Informática y Automática. Univ. Salamanca*, p. 11, [Online]. Available: <http://www.lsi.us.es/redmidas/Capitulos/LMD14.pdf>.
- [27] A. I. González Pineda, “Identificación de factores en la reprobación y deserción mediante técnicas de minería de datos en el área de la energía de la universidad nacional de Loja,” pp. 2–14, 2018.
- [28] D. Beltran, D. Poveda, A. Bolívar, S. Y. Corrales, and H. M. Sarmiento, “Rapidminer,” pp. 1–47, 2010.
- [29] R. R. R. Vega, “Análisis de Sentimientos en Twitter para Descubrir Contenido Xenófobo hacia los Inmigrantes Venezolanos en Ecuador,” 2021.
- [30] H. Merlino, M. Ochoa, E. Diez, and P. Britos, “Gestión Asistida de documentos en una metodología de explotación de información,” *Cent. Ing. Softw. e Ing. del Conoc.*
- [31] R. Timaran and J. Jiménez, “Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM,” *Form. Univ.*, pp. 1–19, 2014.
- [32] V. Galán, “Aplicación de la Metodología Crisp-Dm a un proyecto de Minería de datos en el Entorno Universitario.,” p. 120, 2015.
- [33] C. I. García Osorio, “Metodologías de desarrollo de proyectos de minería de datos - Una visión centrada en CRISP-DM,” *Univ. Burgos*, no. May, 2019, doi: 10.13140/RG.2.2.34208.02566.
- [34] “Reglamento de Régimen Académico. Consejo de Educación Superior,” *Biophys. J.*, pp. 11–14, 2019, doi: 10.1016/S0006-3495(03)74514-7.

- [35] M. Genero, J. Cruz-Lemus, and M. Piattini, "Métodos de investigación en ingeniería del software." pp. 7–9, 2015.
- [36] P. V. Torres-Carrion, C. S. Gonzalez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, "Methodology for systematic literature review applied to engineering and education," *IEEE Glob. Eng. Educ. Conf. EDUCON*, vol. 2018-April, pp. 1364–1373, 2018, doi: 10.1109/EDUCON.2018.8363388.
- [37] X. Pons Diez, M. Gil Lacruz, J. M. Grande Gascon, and M. Marin Jimenez, "Metodo Científico y sus etapas," *Psicol. comunitaria Descr. un caso*, pp. 19–30, 2011, doi: 10.4272/978-84-9745-259-5.ch2.
- [38] Y. Castán, "Introducción al método Científico y Sus Etapas," *Inst. Aragon. Ciencias La Salud*, vol. 2, pp. 1–6, 2006.
- [39] J. D. Velásquez, "Una guía corta para escribir revisiones sistemáticas de literatura. Parte 1," *DYNA*, vol. 82, no. 190, pp. 9–10, 2015, doi: 10.15446/dyna.v82n190.49511.
- [40] J. D. Velásquez, "Una guía corta para escribir revisiones sistemáticas de literatura. Parte 2," *DYNA*, vol. 82, no. 190, pp. 9–12, 2015, doi: 10.15446/dyna.v82n190.49511.
- [41] J. D. Velásquez, "Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura. Parte 3," *DYNA*, vol. 82, no. 190, pp. 9–12, 2015.
- [42] A. Al-Zubidy, J. C. Carver, D. P. Hale, and E. E. Hassler, "Vision for SLR tooling infrastructure: Prioritizing value-added requirements," *Inf. Softw. Technol.*, vol. 91, pp. 72–81, 2017, doi: 10.1016/j.infsof.2017.06.007.
- [43] C. PEREZ LOPEZ and D. SANTIN GONZALEZ, *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Ediciones Paraninfo, S.A, 2007.
- [44] M. De Moya and J. Rodríguez, "La contribución de las reglas de asociación a la minería de datos," *Tecnura*, vol. 13, no. 13, pp. 94–98, 2003.
- [45] D. Carolina and R. Rea, "Detección de patrones de contrabando para la gestión de información de aprehensiones y retenciones utilizando técnicas descriptivas de agrupamiento, asociación y atípicos en minería de datos.," 2021.
- [46] A. Jaramillo Zhingre, "Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje," vol. 28, no. 1, pp. 64–90, 2015.

- [47] J. I. Jairo Villasmil, "Plan De Acción Para El Mejoramiento De Los Talleres Ofrecidos En La Casa Don Bosco De Valencia: Una Propuesta," *Laurus*, vol. 13, no. 25, pp. 263–285, 2007.
- [48] "Formato Plan de Acción," 2017.
- [49] G. Pascal, J. E. Tornillo, C. Minnaard, and M. Comoglio, "Data mining to increase teaching performance in engineering education," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1481, pp. 308–311, 2019, doi: 10.1145/3318396.3318433.
- [50] R. Da Fonseca Silveira, M. Holanda, M. De Carvalho Victorino, and M. Ladeira, "Educational data mining: Analysis of drop out of engineering majors at the UnB - Brazil," *Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019*, pp. 259–262, 2019, doi: 10.1109/ICMLA.2019.00048.
- [51] A. B. Duche-Perez, F. M. Paredes-Quispe, and O. A. Gutierrez-Aguilar, "The Transition from high school to university: Identifying internal and external factors for a successful transition in peruvian students of Architecture and Engineering," *EDUNINE 2019 - 3rd IEEE World Eng. Educ. Conf. Mod. Educ. Paradig. Comput. Eng. Career, Proc.*, pp. 1–6, 2019, doi: 10.1109/EDUNINE.2019.8875751.
- [52] L. Barba-Guamán and P. Valdiviezo-Díaz, "Improve the performance of students in the mathematics learning through Bayesian model," *2017 7th Int. Work. Comput. Sci. Eng. WCSE 2017*, no. Wcse, pp. 349–354, 2017, doi: 10.18178/wcse.2017.06.060.
- [53] S. V. Muñoz-Camacho, T. Gallardo, M. Muñoz-Bravo, and C. A. Muñoz-Bravo, "Probability of student dropout in courses of basic mathematics in professional programs of the Universidad de Los Andes-Venezuela," *Form. Univ.*, vol. 11, no. 4, pp. 33–42, 2018, doi: 10.4067/S0718-50062018000400033.
- [54] C. Russo and A. De Giusti, "Deserción Universitaria En Carreras De Informática De La Unnoba," *Univ. Nac. la Plata*, 2019, [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/79958>.
- [55] L. C. Calvache Fernández, V. Alvarez Vallejo, J. I. Triviño Arbeláez, C. Quiceno Restrepo, and R. Pulgarin Giraldo, "Aplicación de técnicas de minería de datos para la identificación de patrones de deserción estudiantil como apoyo a las estrategias de SARA (Sistema de acompañamiento para el rendimiento académico)," *Octava Conf. Latinoam. sobre el Abandon. en la Educ. Super.*, pp. 1177–1185, 2018, [Online]. Available:

<https://revistas.utp.ac.pa/index.php/clabes/article/view/2021>.

- [56] C. Z. Torres, C. A. Ramos, and J. L. Moraga, "Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año , mediante minería de datos.," *Cienc. Amaz. (Iquitos)*, ISSN 2221-5948, ISSN-e 2222-7431, Vol. 6, N<sup>o</sup>. 1, 2016, págs. 73-84, vol. 6, no. 1, pp. 73–84, 2016, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=5608574>.
- [57] E. R. Ramirez-Saavedra, D. M. Espinosa-Sarmiento, and E. E. Millan-Rojas, "Estrategia para afrontar la deserción universitaria desde las tecnologías de la información y las comunicaciones.," *Rev. Científica*, vol. 1, no. 24, p. 52, 2016, doi: 10.14483/udistrital.jour.rc.2016.24.a5.
- [58] K. B. Eckert and R. Suénaga, "Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos," *Form. Univ.*, vol. 8, no. 5, pp. 3–12, 2015, doi: 10.4067/S0718-50062015000500002.
- [59] INEVAL, "CIE\_ResumenEjecutivoPISA18\_20181123," *Oecd*, p. 24, 2018, [Online]. Available: <http://www.evaluacion.gob.ec/evaluaciones/pisa-documentacion/>.
- [60] I. O. Pappas, M. N. Giannakos, and L. Jaccheri, "Investigating factors influencing students intention to dropout computer science studies," *Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, vol. 11-13-July, no. 7491, pp. 198–203, 2016, doi: 10.1145/2899415.2899455.
- [61] A. Hellas *et al.*, "Predicting academic performance: A systematic literature review," *Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, pp. 175–199, 2018, doi: 10.1145/3293881.3295783.
- [62] L. Chamba-eras, "Ciencias de la Computación basada en evidencia: Guía práctica para desarrollar una Revisión Sistemática de Literatura," pp. 1–8, 2020.
- [63] H. Chanlekha and J. Niramitranon, "Student performance prediction model for early-identification of at-risk students in traditional classroom settings," *MEDES 2018 - 10th Int. Conf. Manag. Digit. Ecosyst.*, pp. 239–245, 2018, doi: 10.1145/3281375.3281403.
- [64] Y. Chen, A. Johri, and H. Rangwala, "Running out of STEM: A comparative study across STEM majors of college students At-Risk of dropping out early," *ACM Int. Conf. Proceeding Ser.*, pp. 270–279, 2018.
- [65] E. M. Trauth, C. C. Cain, K. D. Joshi, L. Kvasny, and K. M. Booth, "The influence

- of gender-ethnic intersectionality on gender stereotypes about IT skills and knowledge,” *Data Base Adv. Inf. Syst.*, vol. 47, no. 3, pp. 9–39, 2016, doi: 10.1145/2980783.2980785.
- [66] O. S. Balogun, S. S. Oyelere, and D. D. Atsa’Am, “Data analytics on performance of computing students,” *ACM Int. Conf. Proceeding Ser.*, pp. 0–1, 2019, doi: 10.1145/3364510.3366152.
- [67] J. P. Salazar-Fernandez, M. Sepúlveda, and J. Munoz-Gama, “Describing educational trajectories of engineering students in individual high-failure rate courses that lead to late dropout,” *CEUR Workshop Proc.*, vol. 2425, pp. 39–48, 2019.
- [68] B. Perez, C. Castellanos, and D. Correal, “Applying Data Mining Techniques to Predict Student Dropout: A Case Study,” *2018 IEEE 1st Colomb. Conf. Appl. Comput. Intell. CoICACI 2018 - Proc.*, pp. 1–6, 2018, doi: 10.1109/CoICACI.2018.8484847.
- [69] M. Mehra, D. R. Kalbande, S. Mankar, and S. Mutsaddi, “Data Mining in Educational Systems for Effective Student Mentoring,” in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, 2019, pp. 1–5, doi: 10.1109/ICAC347590.2019.9036803.
- [70] H. Gulati, “Predictive analytics using data mining technique,” in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015, pp. 713–716.
- [71] O. Sukhbaatar, K. Ogata, and T. Usagawa, “Mining Educational Data to Predict Academic Dropouts: A Case Study in Blended Learning Course,” *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-Octob, no. October, pp. 2205–2208, 2019, doi: 10.1109/TENCON.2018.8650138.
- [72] T. Hasbun, A. Araya, and J. Villalon, “Extracurricular activities as dropout prediction factors in higher education using decision trees,” *Proc. - IEEE 16th Int. Conf. Adv. Learn. Technol. ICALT 2016*, pp. 242–244, 2016, doi: 10.1109/ICALT.2016.66.
- [73] K. J. O. De Santos, A. G. Menezes, A. B. De Carvalho, and C. A. E. Montesco, “Supervised learning in the context of educational data mining to avoid university students dropout,” *Proc. - IEEE 19th Int. Conf. Adv. Learn. Technol. ICALT 2019*, vol. 2161–377X, pp. 207–208, 2019, doi: 10.1109/ICALT.2019.00068.

- [74] G. Tirado-Mendoza, W. Neyra-Lopez, A. Roman-Gonzalez, N. Vargas-Cuentas, D. Llulluy, and U. Lapa-Asto, "Work-in-Progress: Application of academic strategies to strengthen intrinsic motivation and reduce the number of students who abandon engineering programs at universities in Lima-North," *EDUNINE 2019 - 3rd IEEE World Eng. Educ. Conf. Mod. Educ. Paradig. Comput. Eng. Career, Proc.*, pp. 1–4, 2019, doi: 10.1109/EDUNINE.2019.8875844.
- [75] V. Mhetre and M. Nagar, "Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA," *Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017*, vol. 2018-Janua, no. Iccmc, pp. 475–479, 2018, doi: 10.1109/ICCMC.2017.8282735.
- [76] I. Rehman and T. R. Soomro, "Data Mining for Forecasting OGDCL Share Prices Using WEKA," *MACS 2019 - 13th Int. Conf. Math. Actuar. Sci. Comput. Sci. Stat. Proc.*, pp. 1–7, 2019, doi: 10.1109/MACS48846.2019.9024780.
- [77] C. Jalota and R. Agrawal, "Analysis of Educational Data Mining using Classification," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Perspectives Prospect. Com. 2019*, pp. 243–247, 2019, doi: 10.1109/COMITCon.2019.8862214.
- [78] K. I. M. Ramaphosa, T. Zuva, and R. Kwuimi, "Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools," *2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. icABCD 2018*, pp. 1–6, 2018, doi: 10.1109/ICABCD.2018.8465478.
- [79] M. Kaur and A. Girdhar, "Assessment Tool for Outcome Based," *2018 IEEE Int. Conf. Comput. Intell. Comput. Res.*, pp. 1–5.
- [80] S. S. Al-Nadabi and C. Jayakumari, "Predict the selection of mathematics subject for 11th grade students using Data Mining technique," *2019 4th MEC Int. Conf. Big Data Smart City, ICBDS 2019*, pp. 1–4, 2019, doi: 10.1109/ICBDSC.2019.8645594.
- [81] M. Sivasakthi, "Classification and prediction based data mining algorithms to predict students' introductory programming performance," *Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017*, no. Icici, pp. 346–350, 2018, doi: 10.1109/ICICI.2017.8365371.
- [82] S. B. Chandalasetty *et al.*, "Assessing the Relation between Family Background and Juvenile Delinquency using Data Mining," *2019 Int. Conf. Comput. Commun.*

*Informatics, ICCCI 2019*, pp. 1–4, 2019, doi: 10.1109/ICCCI.2019.8822138.

- [83] T. Toivonen and I. Jormanainen, “Evolution of decision tree classifiers in open ended educational data mining,” *ACM Int. Conf. Proceeding Ser.*, pp. 290–296, 2019, doi: 10.1145/3362789.3362880.
- [84] A. Roy, R. R. Morales, E. Fouché, and G. Möhler, “In-database geospatial analytics using python,” *Proc. 2nd ACM SIGSPATIAL Int. Work. Adv. Resilient Intell. Cities, ARIC 2019*, pp. 17–24, 2019, doi: 10.1145/3356395.3365598.
- [85] P. R. Shah, D. B. Vaghela, and P. Sharma, “Faculty performance evaluation based on prediction in distributed data mining,” *ICETECH 2015 - 2015 IEEE Int. Conf. Eng. Technol.*, no. March, pp. 1–5, 2015, doi: 10.1109/ICETECH.2015.7275019.
- [86] G. G. Gokilam and K. Shanthi, “Performance Analysis of Various Data mining Classification Algorithms on Diabetes Heart dataset,” *Compusoft*, vol. 5, no. 3, pp. 2074–2079, 2016.
- [87] R. Sanati-Mehrziy, K. Parkinson, and A. Minaie, “Integration of Data Mining Course in Computer Science Curriculum,” *J. Comput. Sci. Coll.*, vol. 34, no. 2, pp. 87–98, 2018.
- [88] A. K. Varudharajulu and Y. Ma, “Feature-Based Restaurant Customer Reviews Process Model Using Data Mining,” in *Proceedings of the 2018 International Conference on Computing and Big Data*, 2018, pp. 32–37, doi: 10.1145/3277104.3277113.
- [89] I. K. A. Enriko, “Comparative study of heart disease diagnosis using top ten data mining classification algorithms,” *ACM Int. Conf. Proceeding Ser.*, pp. 159–164, 2019, doi: 10.1145/3338188.3338220.
- [90] J. A. Talingdan, “Data mining using clustering algorithm as tool for poverty analysis,” *ACM Int. Conf. Proceeding Ser.*, vol. Part F1479, pp. 56–59, 2019, doi: 10.1145/3316615.3316672.
- [91] A. Xu, Y. Zhou, and H. Chen, “Research on Customer Deposit Order Based on Data Mining Technology,” in *Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City*, 2019, pp. 50–53, doi: 10.1145/3377170.3377233.
- [92] A. K. Varudharajulu and Y. Ma, “Feature-based facebook reviews process model for emanagement using data mining,” *ACM Int. Conf. Proceeding Ser.*, pp. 406–

- 410, 2019, doi: 10.1145/3306500.3306514.
- [93] W. Chango, R. Cerezo, and C. Romero, "Predicting academic performance of university students from multi-sources data in blended learning," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3368691.3368694.
- [94] L. Tan, Y. Chen, L. Lai, and R. Yang, "Research on the Effect of Blended Learning Mode Based on Text Data Analysis," in *Proceedings of the International Workshop on Artificial Intelligence and Education*, 2019, pp. 18–22, doi: 10.1145/3397453.3397460.
- [95] M. J. Domingo, B. D. Gerardo, and R. P. Medina, "Fuzzy Decision Tree for Breast Cancer Prediction," 2019, doi: 10.1145/3373477.3373489.
- [96] S. Houda, A. Naila, and B. Samir, "Machine Learning Based Knowledge Organization in Online Communities," 2019, doi: 10.1145/3372938.3372965.
- [97] I. Wittler, X. Liu, and A. Dong, "Deep learning enabled predicting modeling of mortality of diabetes mellitus patients," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3332186.3333262.
- [98] W. Y. Ayele, "Adapting CRISP-DM for idea mining a data mining process for generating ideas using a textual dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 6, pp. 20–32, 2020, doi: 10.14569/IJACSA.2020.0110603.
- [99] C. Vergaraz, A. Oviedo, and C. Carmona, "Estilos de aprendizaje y minería de datos : un estudio preliminar en el contexto universitario," 2018.
- [100] C. Venkateswara Rao and G. Siva Nageswara Rao, "An effective research on data mining techniques for intrusion detection & learning classes," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 11 Special Issue, pp. 845–849, 2019, doi: 10.35940/ijitee.K1151.09811S19.
- [101] M. Irfan, A. R. Nurhidayat, A. Wahana, D. S. Maylawati, and M. A. Ramdhani, "Comparison of K-Nearest Neighbour and support vector machine for choosing senior high school," *J. Phys. Conf. Ser.*, vol. 1280, no. 2, 2019, doi: 10.1088/1742-6596/1280/2/022026.
- [102] J. R. García-González, P. A. Sánchez-Sánchez, M. Orozco, and S. Obredor, "Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia," *Form. Univ.*, vol. 12, no. 4, pp. 55–62, 2019, doi: 10.4067/S0718-50062019000400055.



- [103] R. T. Pereira, J. C. Zambrano, and A. H. Troya, "Identification of factors associated with academic performance in mathematics in the saber 11th tests applying educational data mining," *Proc. LACCEI Int. Multi-conference Eng. Educ. Technol.*, vol. 2019-July, no. July 2019, pp. 24–26, 2019, doi: 10.18687/LACCEI2019.1.1.297.
- [104] G. Schuh, J. P. Prote, M. Molitor, F. Sauermann, and S. Schmitz, "Databased learning of influencing factors in order specific transition times," *Procedia Manuf.*, vol. 31, pp. 356–362, 2019, doi: 10.1016/j.promfg.2019.03.056.
- [105] A. Vilorio *et al.*, "Determinating student interactions in a virtual learning environment using data mining," *Procedia Comput. Sci.*, vol. 155, no. 2018, pp. 587–592, 2019, doi: 10.1016/j.procs.2019.08.082.
- [106] P. E. C. Chacón, A. Z. González, V. H. M. Domínguez, and P. J. C. Herrera, "Analysis of university faculty performance in the use of MOODLE through data mining techniques: Training needs proposals," *Rev. Educ. a Distancia*, no. 58, 2018, doi: 10.6018/red/58/10.
- [107] B. I. J. Lamarca and S. C. Ambat, "The development of a performance appraisal system using Decision Tree analysis and Fuzzy Logic," *Int. J. Intell. Eng. Syst.*, vol. 11, no. 4, pp. 11–19, 2018, doi: 10.22266/ijies2018.0831.02.
- [108] M. Apaza-luque, J. Manotupa-gil, F. Salas-flores, and J. Sulla, "Parking garage management using classification techniques , fuzzy logic and evolutionary algorithms Gestión de aparcamientos de cocheras utilizando técnicas de clasificación , lógica difusa y algoritmos evolutivos," no. July 2018, pp. 19–21.
- [109] M. Haynes, J. Groen, E. Sturzinger, D. Zhu, J. Shafer, and T. McGee, "Integrating Data Science into a General Education Information Technology Course: An Approach to Developing Data Savvy Undergraduates," in *Proceedings of the 20th Annual SIG Conference on Information Technology Education*, 2019, pp. 183–188, doi: 10.1145/3349266.3351417.
- [110] J. L. C. Ferreira, A. F. Aloise, V. K. Matter, J. L. V. Barbosa, and K. S. F. De Oliveira, "A model for predicting disapproval of apprentices in distance education using decision tree," *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3330204.3330223.
- [111] S. C. K. Tekouabou, W. Cherif, and H. Silkan, "A data modeling approach for classification problems: Application to bank telemarketing prediction," *ACM Int.*

- Conf. Proceeding Ser.*, vol. Part F1481, pp. 1–7, 2019, doi: 10.1145/3320326.3320389.
- [112] J. Sulla-Torres, C. Incalla Nina, M. Rivera Portugal, M. Cossio-Bolanos, and R. Gomez Campos, “Comparison of Classification Algorithms for the Detection of Bone Weakness in Students Using Anthropometric Data,” *Proc. - 2019 Int. Conf. Incl. Technol. Educ. CONTIE 2019*, pp. 56–62, 2019, doi: 10.1109/CONTIE49246.2019.00020.
- [113] G. Ding, “Optimization model of physical education teaching evaluation based on association rule algorithm,” *Proc. - 2019 12th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2019*, pp. 516–519, 2019, doi: 10.1109/ICICTA49267.2019.00116.
- [114] R. A. Lopes, L. A. L. Rodrigues, and J. D. Brancher, “Predicting master’s applicants performance using KDD techniques,” *Iber. Conf. Inf. Syst. Technol. Cist.*, 2017, doi: 10.23919/CISTI.2017.7975752.
- [115] C. R. M. Rosa, M. T. A. Steiner, and P. J. Steiner Neto, “Knowledge discovery in data bases: A case study in a private institution of higher education,” *IEEE Lat. Am. Trans.*, vol. 16, no. 7, pp. 2027–2032, 2018, doi: 10.1109/TLA.2018.8447372.
- [116] B. E. V. Comendador, L. W. Rabago, and B. T. Tanguilig, “An educational model based on Knowledge Discovery in Databases (KDD) to predict learner’s behavior using classification techniques,” *ICSPCC 2016 - IEEE Int. Conf. Signal Process. Commun. Comput. Conf. Proc.*, pp. 1–6, 2016, doi: 10.1109/ICSPCC.2016.7753623.
- [117] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, “BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset,” *IEEE Access*, vol. 8, pp. 29575–29585, 2020, doi: 10.1109/ACCESS.2020.2972627.
- [118] D. Pascual, F. Pla, and S. Sánchez, “Algoritmos de agrupamiento,” *Unpublished*, pp. 164–174, 2007.
- [119] C. García and I. Gómez, “Algoritmos de aprendizaje: knn y kmeans,” *Univ. Carlos III Madrid*, 2006.
- [120] A. C. Álvarez Martínez, “Aplicación De Técnicas De Minería De Datos Para Mejorar El Proceso De Control De Gestión En Entel,” *Univ. Chile Fac. Ciencias Físicas Y Matemáticas Dep. Ing. Ind.*, pp. 1–120, 2012.
- [121] E. R. Sinchi and G. P. Gómez Ceballos, “Acceso y deserción en las universidades.

- Alternativas de financiamiento,” *Alteridad*, vol. 13, no. 2, pp. 274–287, 2018, doi: 10.17163/alt.v14n2.2018.10.
- [122] F. J. Muñoz Rosas and E. Álvarez Verdejo, “Métodos de imputación para el tratamiento de datos faltantes: Aplicación mediante R/Splus,” *Rev. Metod. Cuantitativos para la Econ. y la Empres.*, vol. 7, no. 7, pp. 3–30, 2009.
- [123] G. R. Radío, “Los valores perdidos en el muestreo de poblaciones finitas. Técnicas de imputación,” 2017.
- [124] D. M. Lelly Useshe, “Una introducción a la Imputación de Valores Perdidos,” vol. 11, no. 17, pp. 339–361, 2009.
- [125] I. H. Witten and E. Frank, “Machine Learning Tools and Techniques - WEKA,” no. 2005, 2014.
- [126] A. Dutt, “Clustering Algorithms Applied in Educational Data Mining,” *Int. J. Inf. Electron. Eng.*, no. April, 2015, doi: 10.7763/ijjee.2015.v5.513.
- [127] S. P. Álvarez, “Minería de calidad de datos: Aplicación de técnicas de minería de datos para la evaluación de la calidad de los datos,” pp. 1–98, 2018.

## **11. ANEXOS**

### **Anexo 1. Revisión sistemática de la literatura**

#### **1. Introducción**

Con la finalidad de brindar mayor información y bases para la realización de presente Trabajo de Titulación (TT) se ha realizado una revisión sistemática de literatura (RSL), según como lo manifiesta [36] la RSL es una importante área específica para identificar las preguntas de investigación, así como también para justificar futuras investigaciones en dicha área; dicha temática de interés en el presente TT son las investigaciones sobre los diferentes factores (patrones) que existen en los registros académicos relacionados con asignaturas relacionadas a las matemáticas dentro de ingenierías mediante la aplicación de técnicas de Minería de Datos Educativa, basándose para ello en [39]–[41], [62].

#### **2. Metodología**

Para la RSL se hizo uso de la metodología propuesta por Barbara Kitchenham [36] con el propósito de mostrar resultados válidos y confiables.

##### **2.1. Definir las preguntas de Investigación**

Para un correcto desarrollo de la RSL se plantean directrices que permitirán orientar el proceso de la misma, estas directrices se manifiestan como preguntas de investigación las cuales serán respondidas al culminar la RLS. Se plantean 5 preguntas:

- "¿Cuáles son los factores más comunes presentes en la deserción estudiantil universitaria?",
- "¿Las asignaturas relacionadas con las matemáticas influyen en la deserción académica universitaria?",
- ¿Qué técnicas de Minería de Datos se usan para determinar los factores que influyen en la deserción universitaria?,
- ¿Cuáles son las herramientas que se utilizan para el proceso de minería de datos? Y,
- ¿Cuáles son las principales metodologías para el desarrollo de minería de datos?

## 2.2. Definir las palabras claves

La búsqueda se basará en palabras específicas (inglés, español) que den contestación a las preguntas de investigación, estas son: Minería de datos, deserción, abandono, matemáticas, dropout, student desertion, abandonment, mathematics, maths, Data Mining, influence, factors, engineering students, engineering, higher education, university, tools, education, weka, RapidMiner, Rstudio, Wolfram, Python, Knime, herramientas, metodología, methodology.

## 2.3. Definir los scripts de búsqueda

Para la búsqueda en las fuentes bibliográficas se usarán cadenas de búsqueda que contengan las palabras detalladas en la sección anterior. La Tabla VII muestra las cadenas de búsqueda que se usará para la RSL las mismas que darán contestación a las tres primeras preguntas de investigación.

Tabla VI. Cadenas De Búsqueda Para La RSL.

Fuente Bibliográfica	Cadena de Búsqueda	Cadena de Búsqueda en la Base de Datos	Trabajos Encontrados
ACM	("Data Mining" OR "data mining techniques" OR "educational data mining") AND ("Dropout" OR "desertion") AND engineering students AND Maths AND ("University")	[[All: "data mining"] OR [All: "data mining techniques"] OR [All: "educational data mining"]] AND [[All: "dropout" OR [All: "desertion"]] AND [All: engineering students] AND [All: maths] AND [All: "university"] AND [Publication Date: (01/01/2015 TO 03/31/2020)]	42
Scopus	("Data mining" OR "data mining techniques" OR "educational data mining") AND ("desertion" OR "abandonment") AND "students" AND "Maths" AND ("higher education" OR "University")	( "Data mining" OR "data mining techniques" OR "educational data mining" ) AND ( "desertion" OR "abandonment" ) AND " students" AND "Maths" AND ( " higher education" OR "University" ) AND ( LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2018 ) OR LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) )	6
IEEE	("Data Mining) AND ("dropout" OR "desertion" OR "abandonment") AND " engineering students)	((((( "All Metadata": Data Mining) AND "All Metadata": dropout) OR "All Metadata": desertion) OR "All Metadata": abandonment) AND "All Metadata": engineering students) Filters Applied: 2015 - 2020	48

<b>Scielo</b>	(Minería de datos) AND (Deserción universitaria) AND (Ingeniería)	Expresión: (Minería de datos) AND (Deserción universitaria) AND (Ingeniería)	2
<b>Google Scholar</b>	(Data Mining Techniques or Educational Data Mining) and (dropout or attrition) and (mathematics) and (engineering students) and (University or higher education) and (Year of publication >2105)	(Data Mining Techniques or Educational Data Mining) and (dropout or attrition) and (mathematics) and (engineering students) and (University or higher education) and (Year of publication >2105)	633

Para cumplir con las preguntas de investigación cuatro y cinco, se ha elaborado la Tabla VIII y la Tabla IX, las mismas que contienen las cadenas de búsqueda para conocer las herramientas utilizadas en la minería de datos y, la metodología a utilizar para el desarrollo de la minería de Datos

Tabla VII. Cadenas de búsqueda relacionadas a las herramientas para la MD.

<b>Fuente Bibliográfica</b>	<b>Cadena de Búsqueda</b>	<b>Cadena de Búsqueda en la Base de Datos</b>	<b>Trabajos Encontrados</b>
Scopus	(Education Data Mining) AND (weka OR R OR RapidMiner OR wolfram OR pentaho OR python)	TITLE-ABS-KEY ("data mining" AND (weka OR R OR RapidMiner OR wolfram OR pentaho OR python)) AND DOCTYPE (ar OR cp) AND PUBYEAR > 2019	192
ACM	Data Mining AND (Weka OR RapidMiner OR RStudio OR Python) AND Education	[All: data mining] AND [[All: weka] OR [All: rapidminer] OR [All: rstudio] OR [All: python]] AND [All: education] AND [Publication Date: Past 2 years]	1880
IEEE	Data Mining AND (Weka OR RapidMiner OR RStudio OR Python) AND Education	(((((("All Metadata": Data mining) AND "All Metadata": Weka) OR "All Metadata": RapidMiner ) OR "All Metadata": RStudio) OR "All Metadata": Python) AND "All Metadata": Education)	360

Tabla VIII. Cadena de búsqueda para la metodología a usar en el desarrollo de MD.

Fuente Bibliográfica	Cadena de Búsqueda	Cadena de Búsqueda en la Base de Datos	Trabajos Encontrados
Scopus	"data mining" methodology) AND ("CRISP-DM" OR "SEMMA" OR "KDD") AND ("education"OR"higher education")	ALL ( "data mining" methodology ) AND ( "CRISP-DM" OR "SEMMA" OR "KDD" ) AND ( "education" OR "higher education" ) AND ( LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) OR LIMIT-TO ( PUBYEAR , 2018 ) OR LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) OR LIMIT-TO ( LANGUAGE , "Spanish" ) ) AND ( LIMIT-TO ( ACCESSTYPE(OA) ) )	130
ACM	(Data Mining Methodology) AND (crisp-dm OR semma OR kdd) AND (education)	[All: data mining methodology] AND [[All: crisp-dm] OR [All: semma] OR [All: kdd]] AND [All: education] AND [Publication Date: Past 2 years]	785
IEEE	(Data Mining Methodology) AND (crisp-dm OR semma OR kdd) AND (education)	((((("All Metadata":data mining methodology) AND "All Metadata": CRISP-DM) OR "All Metadata": KDD) OR "All Metadata": SEMMA) AND "All Metadata": education)	55

## 2.4. Definir las bases de datos

Los documentos que serán utilizados en esta revisión bibliográfica serán únicamente los que estén presentes en las fuentes bibliográficas específicas detalladas a continuación, cabe recalcar que el uso de Google Scholar fue utilizado con el propósito de complementar la investigación con la literatura gris:

- ✓ Scopus <https://scopus.com>
- ✓ ACM: <https://dl.acm.org/>

- ✓ Scielo: <http://www.scielo.org>
- ✓ IEEE explore: <https://ieeexplore.ieee.org/>
- ✓ Google Scholar: <https://scholar.google.com/>

## 2.5. Comprobar los scripts de búsqueda

En esta sección se muestra un ejemplo sobre la validación de los scripts en la base de datos IEEE (Ver Figura 25).

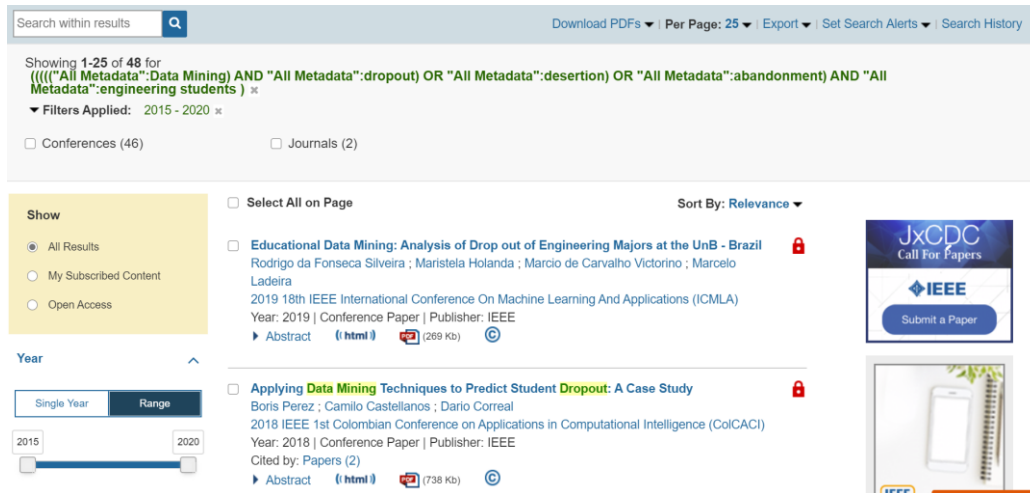


Figura 25. Búsqueda avanzada en la base de datos IEEE.

## 2.6. Definir los criterios de inclusión y exclusión

Buscando desechar la literatura que no ayude significativamente a contestar las preguntas de investigación, se descartará a la misma que no cumpla con:

- **Fecha de publicación:** Máximo 5 años atrás (2015)
- **Motores de búsqueda:** Scopus, ACM, IEEE Explore, Scielo, Google Scholar.
- **Idioma:** La literatura puede estar escrita en español o inglés.
- **Tipos de estudio:** Tesis, artículos científicos, libros.
- **Disciplina:** Educación Superior

## 2.7. Selección de artículos

Luego de establecer los lineamientos para ejecutar la RSL, se procede a realizar la búsqueda de literatura en las fuentes bibliográficas establecidas, aplicando las cadenas de búsqueda correspondientes a cada buscador. Al



culminar con la búsqueda de literatura, la aplicación de los criterios de exclusión e inclusión se verifica un total de 21 trabajos relacionados que responden a las preguntas de investigación 1, 2 y 3 de la presente RSL, los trabajos relacionados se observan en la Tabla 2; La Tabla 5 muestra los 25 trabajos relacionados acerca de la herramienta más utilizada para el proceso de Minería de Datos; Mientras que, la Tabla 6 presenta como resultado un total de 23 trabajos relacionados acerca de la búsqueda sobre la metodología de minería de datos a utilizar para el desarrollo del TT.

## 2.8. Ejecución de los resultados

Una vez establecidos necesario para la ejecución de la RSL se procede a su ejecución, aplicando todos los criterios y procedimientos establecidos anteriormente.

### 2.8.1. Extracción de Información

La extracción de la información se realizó en base los criterios de inclusión y exclusión descritos en la sección 2.6; considerar que los trabajos relacionados son en base a los factores que provocan el abandono estudiantil universitarios relacionando con las matemáticas, así como también la técnica de Minería de Datos Educativa utilizada para el desarrollo del trabajo.

Tabla IX. Trabajos relacionados obtenidos de la RSL

Código	Título	Año de publicación	Causa de Abandono	Técnica utilizada
TR01	Investigating Factors Influencing Students' Intention to Dropout Computer Science Studies [60]	2016	Académicos	--
Aporte	Se trata de una investigación dentro del área de educación en informática (CS) en la cual se ha centrado en identificar las razones por las cuales los estudiantes no terminan sus estudios en CS. Este trabajo tiene como objetivo examinar empíricamente cómo la intención de los estudiantes de completar sus estudios (retención) en CS se ve afectada por las variables que juegan un papel clave en la educación superior. Para hacerlo, se identificaron las siguientes variables: año de estudios, género, edad, esfuerzo de los estudiantes, ausencia de clases, promedio de calificaciones			

	esperado (GPA) y promedio actual, y prueban su efecto sobre la retención, basado en las respuestas recopiladas de 241 estudiantes de CS. El año de estudios y el esfuerzo tienen efectos positivos en la intención de los estudiantes de terminar sus estudios en CS. El GPA tiene un efecto negativo en las intenciones de los estudiantes de terminar sus estudios.			
<b>TR02</b>	Modelo de predicción del rendimiento del alumno para la identificación temprana de alumnos en riesgo en entornos de aula tradicionales [63]	2018	Académicos	Redes Neuronales
<b>Aporte</b>	En esta investigación, se enfoca en desarrollar modelos de predicción en un entorno con falta de información para identificar estudiantes en riesgo que tienen tendencia a recibir bajos los grados. Para construir los modelos, cinco técnicas de aprendizaje automático fueron utilizadas para construir los modelos. Los atributos utilizados en este trabajo eran la información mínima que la Universidad tenía típicamente almacenado para cada sujeto. Aunque las precisiones generales de predicción en todas las categorías fueron bajas, el desempeño en la identificación de estudiantes en riesgo fue bastante prometedor.			
<b>TR03</b>	Predicting academic performance: a systematic literature review [61]	2018	Académicos, Sociales, (Demográficos, Familiares, etc.)	Árboles de decisión
<b>Aporte</b>	Esta investigación busca identificar características que puedan usarse para hacer predicciones, identificar algoritmos que puedan mejorar las predicciones y cuantificar aspectos del rendimiento de los estudiantes. Además, la investigación para predecir el rendimiento del estudiante busca determinar características interrelacionadas e identificar las razones subyacentes por las cuales ciertas características funcionan mejor que otras. Este informe presenta una revisión sistemática de la literatura del trabajo en el área de predicción del rendimiento de los estudiantes.			
<b>TR04</b>	Mejorar el rendimiento de los estudiantes en el aprendizaje de las matemáticas a través del modelo bayesiano [52].	2017	Académicos	Modelo Bayesiano
<b>Aporte</b>	Esta investigación muestra el uso de técnicas de minería de datos y software de herramientas matemáticas como un recurso educativo complementario en la comprensión de las matemáticas. El proceso de rendimiento de las matemáticas en el primer y tercer nivel de Ciencias de la Computación y Electrónica ha utilizado estas técnicas de extracción de datos y herramientas matemáticas como parte de un proceso de			

	<p>formación. Las matemáticas son fundamentales en el proceso de formación de la educación de los futuros profesionales. Los estudiantes de grupos experimentales utilizaron el software Wolfram Mathematica. El modelo bayesiano mostró la predicción de la tasa de aprobación de los estudiantes. Parte de la experiencia en esta investigación también fue obtener la percepción de los estudiantes a través de una encuesta. Como resultado, determinamos la necesidad de este paquete de matemáticas como un recurso educativo complementario, que apoya la capacidad de cálculo entre otros.</p>			
<b>TR05</b>	<p>Quedarse sin STEM: un estudio comparativo entre estudiantes de STEM de estudiantes universitarios en riesgo de abandonar temprano [64]</p>	2018	Académicos	Regresión Logística, Árboles de decisión.
<b>Aporte</b>	<p>Las instituciones de educación superior en los Estados Unidos y en todo el mundo occidental tienen un problema crítico de desgaste de los estudiantes universitarios y este problema es grave en los campos de Ciencia, Tecnología, Ingeniería y Matemáticas (STEM). Los estudiantes son especialmente vulnerables en los años iniciales de sus programas académicos; Más del 60% de los abandonos ocurridos en los primeros dos años. Por lo tanto, la identificación temprana de los estudiantes en riesgo es crucial para una intervención enfocada si las instituciones deben apoyar a los estudiantes hacia su finalización. En este documento se desarrolla y evalúa un marco de análisis de supervivencia para la identificación temprana de estudiantes en riesgo de abandono. Se compara el rendimiento de los métodos de análisis de supervivencia con otros enfoques de aprendizaje automático, incluida la regresión logística, los árboles de decisión y el impulso. Los métodos propuestos determinan un buen rendimiento para la predicción temprana de estudiantes en riesgo y también pueden predecir cuándo un estudiante abandonará la escuela con alta precisión. Además, se realizó un análisis comparativo de nueve especializaciones diferentes con diferentes niveles de rigor académico, desafío y alumnado. Este estudio permite a los asesores y administradores universitarios intervenir con anticipación para mejorar la retención de estudiantes.</p>			
<b>TR06</b>	<p>La influencia de la interseccionalidad de género y etnia en los estereotipos de género sobre las habilidades y el conocimiento de TI [65].</p>	2016	Académicos	--
	<p>Esta investigación busca intentar comprender mejor el desequilibrio de género en el campo de la tecnología de la información (TI) examinando los</p>			

	<p>estereotipos de género sobre las habilidades y el conocimiento en la profesión de TI. Se realizó una encuesta a 4046 estudiantes universitarios en los Estados Unidos para examinar los estereotipos de género de los estudiantes universitarios contemporáneos (hombres y mujeres blancos, negros y latinos) sobre las habilidades y el conocimiento en la profesión de TI. La Teoría de las diferencias individuales de género y TI se utilizó como la teoría motivadora para este estudio porque permitió la incorporación de la interseccionalidad género-étnica en el diseño de la investigación. Los resultados revelaron que, si bien existen los estereotipos de género sobre las habilidades y el conocimiento involucrados en la profesión de TI, no son uniformes en todos los miembros de un grupo de género. Los hombres tendieron a calificar todas las habilidades como más masculinas que las mujeres encuestadas. Las habilidades técnicas fueron estereotipadas más consistentemente por hombres y mujeres en cada uno de los grupos étnicos de género que las habilidades no técnicas. Sin embargo, los estereotipos de género sobre las habilidades no técnicas fueron más controvertidos y revelaron variaciones tanto dentro del género como dentro de la etnia. La calificación de las estudiantes no técnicas de las habilidades no técnicas como menos masculinas que los hombres sugiere que estas habilidades no técnicas se están incorporando en el sentido de identidad de género de las mujeres. Estos resultados muestran que la interseccionalidad género-étnica proporciona una explicación importante para la variación dentro del género en los estereotipos de género que tienen los estudiantes universitarios contemporáneos.</p>			
<b>TR07</b>	Data analytics on performance of computing students [66].	2019	Académicos	Regresión Lineal
<b>Aporte</b>	<p>Este estudio reveló que existen fuertes relaciones lineales entre los GPA y CGPA de los estudiantes de computación sobre el período de este estudio. Todos los resultados de los modelos de regresión se ajustan significativamente a los datos utilizados y cada una de las regresiones. Los parámetros contribuyen a la importancia del modelo. El maestro podría usar el resultado de este estudio para comprender las tendencias en el rendimiento académico del estudiante de informática para proporcionar el apoyo necesario en cada etapa de los estudiantes durante su progreso académico. Además, los resultados de este estudio servirán como motivación para los maestros y buscar signos de problemas en cada estudiante También ayudará a mantener un seguimiento registro de cada alumno prediciendo su rendimiento académico rendimiento para orientar, aconsejar y nutrir a los estudiantes. El resultado de este estudio fomentará la competencia entre los estudiantes y motiva a los estudiantes a trabajar más duro para mejorar su desempeño para tener éxito.</p>			
<b>TR08</b>	Data Mining to Increase Teaching Performance in	2019	Académicos, Sociales	Redes Neuronales

	Engineering Education [49].			
<b>Aporte</b>	El objetivo principal de este trabajo es presentar los resultados de la aplicación de técnicas de minería de datos dirigidas a medir el desempeño docente en la Educación en Ingeniería. La unidad de estudio estará compuesta por tres conjuntos de conocimientos: Ciencias Básicas, Tecnologías Básicas y Tecnologías Aplicadas de las Carreras de Ingeniería. Los indicadores son métricas sobre el rendimiento y el comportamiento del alumno. Aplicamos el análisis de componentes principales (PCA) para encontrar factores clave que describen el conjunto de conocimiento definido. Finalmente, establecemos cinco criterios para explicar cada uno. Estos criterios son articulación vertical, estrategias de retención, metodologías de enseñanza, metodologías de evaluación y métodos y técnicas de aprendizaje.			
<b>TR09</b>	Probability of student dropout in courses of basic mathematics in professional programs of the Universidad de Los Andes-Venezuela [53].	2018	Académicos, Sociales	--
<b>Aporte</b>	El objetivo del presente trabajo es describir las probabilidades de deserción estudiantil en cursos de matemáticas básicas para diferentes programas profesionales de la Universidad de Los Andes-Venezuela. Este trabajo es el resultado de una investigación realizada en el período de 2011 a 2016, que asumió un diseño de investigación cuantitativo y siguió la alternativa metodológica del estudio de caso único. Se aplicaron diferentes métodos del nivel teórico, empírico y estadístico-matemático. Para tal efecto, se construyó un modelo Logit que permitiera evaluar si la probabilidad de deserción puede ser predicha considerando características personales (como género) y aspectos institucionales, como el tipo de periodo académico, el tipo de curso académico y el número de estudiantes por curso. Los resultados obtenidos evidencian de manera relevante, la mayor probabilidad de desertar de un periodo académico regular, libre escolaridad, cursos de mayor de cincuenta estudiantes y género femenino.			
<b>TR10</b>	Educational Data Mining: Analysis of Drop out of Engineering Majors at the UnB – Brazil [50].	2019	Académicos	Modelo lineal generalizado (GLM), Algoritmo de refuerzo (GBM) Bosque aleatorio (RF).
	Este artículo presenta un análisis de datos sobre la deserción de estudiantes universitarios de ingeniería en la Universidad de Brasilia (UnB), Brasil. Lo que se pretende es dar respuesta a la pregunta de investigación: ¿Cuáles son los principales factores que motivan a los estudiantes de ingeniería a abandonar las especialidades de ingeniería en la UnB? Para			

	<p>ello se recopiló datos sociales y de rendimiento de los estudiantes de ingeniería de 2009 a 2019. Algunos de los datos pueden considerarse raros en estudios similares, como la distancia de los estudiantes desde el hogar al campus y factores como los de los estudiantes solicitudes de licencia en lugar de factores de rendimiento. Utilizamos tres técnicas de minería de datos: modelo lineal generalizado (GLM), algoritmo de refuerzo (GBM) y bosque aleatorio (RF). Los resultados del estudio mostraron que los estudiantes internacionales merecen cierta atención de la Universidad y cursos como Física 1 pueden ser un desafío para los estudiantes de ingeniería.</p>			
<b>TR11</b>	<p>Influence of Student Diversity on Educational Trajectories in Engineering High-Failure Rate Courses that Lead to Late Dropout [67].</p>	2019	Académicos, Sociales	Modelo de Inversión
<b>Aporte</b>	<p>Este estudio de caso busca responder cómo las habilidades de género, ingresos y matemática de entrada pueden explicar las diferencias en las trayectorias educativas de los estudiantes de ingeniería en los cursos de alta tasa de reprobación que conducen al abandono tardío. Expedientes académicos de 794 estudiantes de ingeniería en la Universidad Austral de Chile que pertenecen a las cohortes 2007 a 2009, fueron extraídos y analizados usando técnicas de descubrimiento de minería de procesos. Se crearon modelos de trayectorias educativas en cursos de alta tasa de fracaso y luego se analizaron utilizando el Modelo de Inversión como marco de referencia. Los resultados revelan lo siguiente: el abandono tardío está relacionado con el número de semestres consecutivos que un estudiante mantiene pendientes cursos reprobados; los estudiantes de bajos ingresos y aquellos con habilidades matemáticas bajas de entrada tienden a ser más persistentes, incluso si tienen trayectorias insatisfactorias; Las estudiantes tienden a ser más reacias al riesgo cuando tienen resultados insatisfactorios. Comprender las trayectorias educativas de los estudiantes que terminan en una deserción tardía puede ayudar a los gerentes y a los responsables de las políticas a mejorar el plan de estudios, las condiciones de ingreso y los programas que apoyan a los estudiantes desfavorecidos.</p>			
<b>TR12</b>	<p>Applying Data Mining Techniques to Predict Student Dropout: A Case Study [68].</p>	2018	Académicos	Árboles de decisión
<b>Aporte</b>	<p>Este artículo muestra los resultados preliminares para predecir la deserción de estudiantes a partir de un gran conjunto de datos heterogéneos de datos demográficos de estudiantes y registros de transcripciones. En los hallazgos se descubrió que el rendimiento de los cursos de ingeniería de sistemas es correlacionado con el rendimiento de los cursos de física y matemáticas. La irregularidad (desviación estándar de los promedios del</p>			

	<p>término) es positivamente correlacionado con la deserción. Además, los resultados experimentales mostraron que la mejor área bajo la curva fue lograda por el modelo de árbol de decisión (0.94), por lo que esta precisión podría tener la confianza suficiente para ayudar a abandonar temprano detección temprana. Cuatro características eran necesarias (SEMESTRE, PROM SIS, VECES ISIS, PROM ACUM) para lograr esta exactitud. Implica que los cursos relacionados con SE tienen el mayor impacto en la predicción de abandono.</p>			
<b>TR13</b>	Data Mining in Educational Systems for Effective Student Mentoring [69].	2019	Académicos, Sociales	Máquina de vectores de soporte
<b>Aporte</b>	<p>Este documento habla de la integración de un programa informático impulsado por Machine Learning, con tutoría para garantizar un sistema de orientación eficaz. En el sistema propuesto, se utilizó Educational Data Mining (EDM) para crear un modelo de predicción basado en la regresión para ayudar a los profesores en el proceso de tutoría mediante la generación de una estimación del rendimiento futuro de los estudiantes. Para hacer este sistema más preciso, se utilizaron factores aberrantes que afectan el rendimiento de los estudiantes, como el tiempo de viaje, el tipo de residencia, el género, la casta, la educación de los padres y el estado laboral, junto con características académicas y analíticas convencionales como CGPA y extracurricular para entrenar el modelo de aprendizaje automático. Después de la implementación, se observó que el modelo de regresión construido usando Máquinas de vectores de soporte (con Kernel RBF) tenía el menor error cuadrado. Este modelo se integró luego con un sistema de tutoría para crear una aplicación informática para ayudar a los profesores. Con la ayuda de la aplicación, los mentores pueden crear mejor un plan a medida para cada alumno y mejorar aún más la efectividad de la tutoría del alumno.</p>			
<b>TR14</b>	Análisis predictivo utilizando la técnica de minería de datos [70].	2015	Académicos	Árboles de decisión
<b>Aporte</b>	<p>El presente documento indica que existe conocimiento oculto en los sistemas de datos educativos y puede ser procesado mediante la tecnología de minería de datos para analizar los factores que afectan el abandono de los estudiantes, así como también generar información valiosa para la toma de decisiones de los interesados para mejorar la calidad del sistema de educación superior. En este seminario se ha usado datos reales de un centro de estudios de la Universidad Nacional Abierta Indira Gandhi. Se ha recopilado datos de varias fuentes, como bases de datos universitarias, formularios de encuestas, etc. Se aplican varios pasos de minería para deducir resultados útiles. Se compararon varios escenarios y se calculó la precisión. Este estudio presenta el trabajo de minería de datos para predecir la función de abandono de los estudiantes. Este artículo presenta el análisis del conjunto de datos utilizando algoritmos de minería</p>			



	de datos. Después del análisis, el resultado serán los factores principales que más afectan a los estudiantes que abandonan los cursos abiertos (tasa de abandono). Antes de aplicar algoritmos de clasificación, también se utilizan algunos algoritmos de selección de características para obtener resultados de predicción refinados. Tal información de análisis y predicción ayudará a la administración de la Universidad y a los maestros a realizar los cambios necesarios para impartir una mejor educación. La herramienta utilizada para la selección de características y la minería es weka.			
<b>TR15</b>	Mining Educational Data to Predict Academic Dropouts: A Case Study in Blended Learning Course [71].	<b>2019</b>	Académicos	Arboles de decisión
<b>Aporte</b>	Presenta una propuesta sobre un esquema de predicción simple utilizando el análisis del árbol de decisión con el propósito de clasificar para identificar a los estudiantes propensos a abandonar la escuela a mediados del semestre en función de las características del curso del año anterior para ese curso. Los datos incluyeron las actividades en línea de 717 estudiantes en cursos obligatorios de segundo año con estilos de aprendizaje combinados, el 79% de los estudiantes que abandonaron la escuela se predijeron correctamente.			
<b>TR16</b>	Actividades extracurriculares como factores de predicción de abandono en la educación superior utilizando árboles de decisión [72].	2016	Sociales	Árboles de decisión
<b>Aporte</b>	Este artículo estudia la importancia de las actividades extracurriculares para predecir el abandono en los estudiantes de dos títulos de Bachiller en Ciencias (Ingeniería y Negocios). Se recopilieron datos de 4.840 estudiantes y se capacitaron y validaron dos modelos, uno que incluía todos los datos y otro que eliminaba créditos por valor de cursos, lo que demuestra que las actividades extracurriculares son excelentes predictores de abandono.			
<b>TR17</b>	Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout [73].	2019	Académicos	Árboles de Decisión, vecino cercano, redes neuronales, bosques aleatorios.
<b>Aporte</b>	En este artículo, se propone un enfoque computacional que utiliza la minería de datos educativos y diferentes técnicas de aprendizaje supervisado (árboles de decisión, vecino K más cercano, redes neuronales, máquinas de vectores de soporte, naive bayes y bosques aleatorios) para			



	<p>evaluar el comportamiento de diferentes modelos de predicción para Identificar el perfil de estudiantes universitarios en riesgo en un entorno universitario brasileño. Este artículo presentó el análisis de diferentes técnicas de aprendizaje supervisado en el contexto de la minería de datos educativa para evitar la deserción escolar de los estudiantes universitarios. Fue encontrado que los estudiantes abandonaron sus respectivos programas más frecuentemente en el cuarto semestre para los programas Ciencias de la Computación y Sistemas de Información, y en el sexto semestre para el Ciencias de la Computación los algoritmos tuvieron sus mejores resultados ante los niveles de deserción. Mientras que algunos algoritmos presentaron resultados sólidos para cada semestre (RF y árboles de decisión), algunos de ellos no pudieron para lograr resultados tan altos.</p>			
<b>TR18</b>	<p>The Transition from high school to university: identifying internal and external factors for a successful transition in peruvian students of Architecture and Engineering [51].</p>	2019	Académicos	--
	<p>El objetivo de esta investigación es determinar qué factores internos y externos permiten a los nuevos estudiantes con éxito adaptarse al entorno universitario. Estudiantes de primer año de la Universidad Católica de Santa María (Arequipa, Perú) fueron los participantes de esta investigación. Estudiantes de industrial Ingeniería (81), Ingeniería de minas (22), Ingeniería en Sistemas (18) y Arquitectura (47). Para recopilar los datos se elaboró un cuestionario con 10 preguntas (tipo cerrado y de una opción preguntas) y se utilizó un grupo focal con 32 estudiantes. Los resultados revelaron que 1 de cada 3 estudiantes tuvo varias dificultades para lograr una transición exitosa. Además, presentaron un bajo rendimiento en uno o más cursos al final de su primer año de estudio. Cuatro factores generales identificados habilitados ellos para adaptarse e integrarse en el entorno universitario exitosamente: apoyo familiar recibido durante el primer semestre de estudios, competencias adquiridas en la escuela secundaria, la desempeño como estudiante en la Universidad y el objetivo propuesta por el estudiante universitario durante el primer semestre de estudios.</p>			
<b>TR19</b>	<p>Work-in-Progress: Application of academic strategies to strengthen intrinsic motivation and reduce the number of students who abandon engineering programs</p>	2019	Académicas	--

	at universities in Lima – North [74].			
	<p>En este trabajo se demostró cuantitativamente que durante un semestre completo de aplicación de la academia estrategias, ha sido posible reducir en el semestre 2017-II y 2018-I el número de estudiantes que abandonan la ingeniería programa. Es importante señalar que, para el éxito de la motivación intrínseca en los alumnos, complementaria y se consideran escenarios transversales tales como: feria de proyectos, concursos de proyectos y congresos de investigación dentro de la Universidad en sí. La diferencia con otras estrategias es que hemos considerado características por niveles, analizando la condición en la que los estudiantes de la institución se basan en el rango de edad, la característica de motivación, la disponibilidad de tiempo y la situación en la que están al mismo nivel académico y personal.</p>			
<b>TR20</b>	<p>Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos [58].</p>	2015	Académicas	Árboles de decisión
	<p>Se analiza información académica con el objetivo de identificar factores que influyen sobre la deserción de los estudiantes de la Carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina, mediante la aplicación de una técnica de minería de datos. La fuente de datos contiene información proporcionada al ingreso (personales y antecedentes educativos) y la que se genera durante el periodo de estudios. Se realiza la selección y depuración de datos, utilizando diferentes criterios de representación y aplicación de algoritmos de clasificación como árboles de decisión, redes bayesianas y reglas. Se identifica como variables influyentes en la deserción, asignaturas aprobadas, cantidad y resultado de asignaturas cursadas, procedencia y edad de ingreso del estudiante. Mediante este proceso fue posible identificar los atributos que caracterizan a los casos de deserción y su relación con el desempeño académico, especialmente en el primer año de la carrera.</p>			
<b>TR21</b>	<p>Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos [56].</p>	2016	Académicos	Random Forest
	<p>El objetivo de esta investigación fue analizar la deserción de los estudiantes mediante técnicas de minería de datos y obtener un modelo que fuese capaz de clasificar estudiantes desertores a partir de los datos socioeconómicos y académicos de los estudiantes de Carreras de pregrado en la Universidad Arturo Prat, Chile. Para el desarrollo de este proyecto se</p>			

	<p>utilizó CRISP-DM como metodología para guiar las etapas del proyecto y se analizaron tres diferentes modelos de clasificación: árboles de decisión, métodos bayesianos y redes neuronales, con el fin de evaluar su comportamiento, encontrándose que Random Forest es el algoritmo de mejor desempeño general, con un 88,9% de exactitud, mientras que el algoritmo Naive Bayes resulto ser el más adecuado para dar respuesta a los objetivos del negocio, dados los niveles de sensibilidad alcanzados. Mediante los experimentos realizados se determinó que las variables académicas de ingreso de los estudiantes no resultan significativas para explicar la deserción de primer año.</p>
--	---

Para la extracción de información acerca de la herramienta para realizar Minería de Datos Educativa, tenemos la Tabla XI:

Tabla X. Trabajos relacionados con la herramienta a utilizar

<b>Trabajos Relacionados en base a las Herramientas Utilizadas</b>			
<b>Código</b>	<b>Título</b>	<b>Año de publicación</b>	<b>Herramienta utilizada</b>
TR01	Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA [75].	2018	WEKA
<b>Aporte</b>	Este artículo se centra en identificar a los estudiantes lentos, medios y rápidos entre los estudiantes y mostrarlo mediante un modelo de minería de datos predictivo mediante algoritmos basados en la clasificación. Los detalles del estudiante han sido referidos del Departamento MCA del Instituto Sardar Patel de Tecnología y la predicción de los estudiantes se realiza mediante la aplicación de Naive Bayes, J48, ZeroR y Random Tree usando WEKA como una herramienta de código abierto.		
TR02	Data Mining for Forecasting OGDCL Share Prices Using WEKA [76].	2019	WEKA
<b>Aporte</b>	El objetivo de este documento es utilizar la herramienta WEKA para el análisis de datos de los precios de las acciones de OGDCL. Se han utilizado dos algoritmos de predicción, es decir, SMOreg y Perceptrón Multicapa. El precio abierto, alto, bajo y de cierre de la acción de OGDC se ha pronosticado durante diez días. Se han comparado los resultados obtenidos de ambos algoritmos. Se encontró que SMOreg proporciona el resultado más preciso de los dos algoritmos para el conjunto de datos dado.		
TR03	Analysis of Educational Data Mining using Classification [77].	2019	WEKA
<b>Aporte</b>	En este artículo se discute acerca de las distintas técnicas de minería de datos que son útiles para predecir el nivel de rendimiento de los estudiantes. Para ello, se hace uso del conjunto de datos de kalboard 360 el cual lo aplica en el software weka para analizar las técnicas de minería de datos.		

TR04	Predictive analytics using data mining technique [70].	2015	WEKA
<b>Aporte</b>	Este estudio presenta el trabajo de minería de datos para predecir la característica de abandono de los estudiantes. Este documento presenta el análisis del conjunto de datos utilizando algoritmos de minería de datos. Después del análisis, el resultado serán los principales factores que afectarán más a la deserción de los estudiantes de los cursos abiertos (tasa de abandono). Antes de aplicar algoritmos de clasificación, también se utilizan algunos algoritmos de selección de características para obtener resultados de predicción refinados. La herramienta utilizada para la selección de características y la minería es weka.		
TR05	Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools [78].	2018	WEKA
<b>Aporte</b>	utilizada para mejorar la calidad de la educación. El conjunto de datos de los registros académicos de los alumnos se aplicó en los algoritmos de clasificación Naive Bayes, BayesNet, JRip y J48 utilizando la herramienta Weka. Los alumnos se clasificaron según su ciudad, escuela, calificaciones y resultados de matemáticas. Se cubrieron aproximadamente 678 datos de alumnos. Este artículo muestra la comparación de cuatro clasificadores y encuentra el algoritmo de clasificación de mejor desempeño entre todos. En base a los resultados obtenidos, encontramos que el algoritmo J48 superó a los otros algoritmos con una precisión de predicción del 99.13%.		
TR06	A Framework for the Indirect Assessment Tool for Outcome Based Education Using Data Mining [79].	2019	WEKA
<b>Aporte</b>	Este documento es un intento de proporcionar el marco para automatizar el proceso de logro utilizando los datos de evaluación y mapeo de los resultados del programa (PO) y los resultados del curso (CO) recuperados de las herramientas de evaluación indirecta. Hay varios tipos de encuestas como encuesta de ex alumnos, encuesta integrada de cursos y encuesta industrial. Se lleva a cabo una encuesta para verificar el nivel de satisfacción del estudiante con respecto al proceso de aprendizaje de la enseñanza. Las preguntas de la encuesta se están mapeando con los CO y posteriormente con los PO. Luego, el maestro hace diferentes preguntas a los alumnos. De esta manera se obtienen datos para cálculos posteriores. Aquí, después, con la ayuda de los cálculos, se calculan los niveles de logro de los CO que se utilizan además en los cálculos de los PO. A continuación, al final se obtiene el logro de diferentes PO. Además, el logro calculado se utiliza para la predicción utilizando la herramienta Weka y corresponde a que las diferentes clases están asociadas a esos datos. De esta manera, se puede mejorar el rendimiento de los estudiantes mirando los diferentes valores en todos los cursos y por lo tanto conduce al desarrollo general de los estudiantes. Finalmente, los valores que se obtienen en la tabla de minería de datos, se pueden utilizar como un paso más para diferentes sugerencias que conduce al desarrollo general de OBE.		

TR07	Predict the selection of mathematics subject for 11th grade students using Data Mining technique [80].	2019	WEKA
<b>Aporte</b>	El objetivo principal de este proyecto es predecir la selección de asignaturas de matemáticas para estudiantes de 11º grado utilizando la técnica de minería de datos. Antes de la creación del modelo en este proyecto, los datos deben ser pre procesados después de la extracción del portal educativo, como la limpieza y reducir los errores y el ruido. Esta investigación utiliza los datos de 2000 estudiantes como un caso de estudio para estudiar en este proyecto. Los datos extraídos de uno del sector educativo de Omán. Se aplicaron diferentes datos de educación (EDM), utilizando métodos de clasificación. El algoritmo utilizado en este proyecto para predecir la selección de estudiantes fue J48, muestra logística y Naive Bayes. Los resultados de este proyecto muestran que muchos estudiantes deben elegir las matemáticas aplicadas debido a su nivel en materia de matemáticas.		
TR08	Classification and prediction based data mining algorithms to predict students' introductory programming performance [81].	2017	WEKA
<b>Aporte</b>	Este documento se enfoca en predecir el rendimiento de la programación introductoria de los estudiantes de primer año de licenciatura en el curso de Aplicación de Computadora mediante un modelo de minería de datos predictivo utilizando algoritmos basados en clasificación. Los datos recopilados contienen los datos demográficos de los estudiantes, la calificación en la programación introductoria en la Universidad y la calificación en la programación introductoria en el examen que contiene 60 preguntas. Los datos recopilados se aplicaron en varios algoritmos de clasificación, como Percepción multicapa, Naïve Bayes, SMO, J48 y REPTree utilizando WEKA. Como resultado, las estadísticas se generan en función de todos los algoritmos de clasificación y también se realiza la comparación de los cinco clasificadores para predecir la precisión y encontrar el algoritmo de clasificación con mejor rendimiento entre todos.		
TR09	Assessing the Relation between Family Background and Juvenile Delinquency using Data Mining [82].	2019	WEKA
<b>Aporte</b>	En este documento, utilizamos la técnica de minería de reglas de asociación de minería de datos para realizar análisis sobre el factor de riesgo de antecedentes familiares en la delincuencia juvenil con ayuda del conjunto de datos de delitos juveniles de la India. Los resultados del experimento se investigan para evaluar la relación entre los antecedentes familiares y la delincuencia juvenil en la India. Para la minería de datos se ha utilizado la herramienta weka.		
TR10	Evolution of Decision Tree Classifiers in Open Ended Educational Data Mining [83].	2019	WEKA, RAPIDMINER

<b>Aporte</b>	El estudio indica que, con el tiempo, los modelos se generalizan mejor y el método AUI ayuda a evitar el problema del sobreajuste. Además, el estudio indica que la naturaleza cíclica del método AUI facilita una generación de conocimiento más profunda a partir del conjunto de datos, si el usuario final conoce el contexto. El modelo predictivo ajustable afecta a la percepción del usuario final y los efectos de ajuste a la salida del modelo predictivo. Cuando se aplica en ciclos, el método AUI genera nuevos conocimientos desde el contexto educativo.		
TR11	In-Database Geospatial Analytics using Python [84].	2019	PYTHON
<b>Aporte</b>	Propone un marco implementado en Python, que integra el análisis geoespacial en una base de datos espacial (es decir, IBM DB2). El marco traduce internamente las funciones espaciales escritas por el usuario en consultas SQL, que siguen los estándares del Open Geospatial Consortium (OGC) y pueden operar en geometrías simples y múltiples. Luego demostramos cómo combinar los resultados de las operaciones espaciales con métodos de visualización como los mapas coropléticos dentro de los cuadernos Jupyter.		
TR12	Faculty performance evaluation based on prediction in distributed data mining [85].	2015	WEKA
<b>Aporte</b>	Proporciona una mejor solución para el problema de predecir y analizar el rendimiento del profesorado en la minería de datos distribuida. Con el uso de la minería de datos distribuida podemos obtener datos de las diferentes fuentes y luego podemos aplicar el algoritmo de clasificación. Mediante la clasificación, podemos obtener una mayor eficiencia y precisión en la medición del desempeño del profesorado. Y podemos construir el modelo de predicción de desempeño basado en las habilidades, puntualidad y desempeño de la Facultad en varias pruebas. Esta técnica de clasificación se prueba en la herramienta WEKA para obtener resultados precisos.		
TR13	Performance Analysis of Various Data Mining Classification Algorithms on Diabetes Heart dataset [86].	2016	WEKA
<b>Aporte</b>	Estas técnicas de minería de datos se aplican en la creación de software para modelos de clasificación rápidos y fáciles. La identificación temprana tiene módulos de alto riesgo que también pueden tener una gran cantidad de fallas. Los modelos de árbol de clasificación son simples y efectivos como modelos de predicción de calidad de software, mientras que las predicciones de defectos de dichos modelos se pueden utilizar para lograr una alta confiabilidad de software .		
TR14	Integration of data mining course in computer science curriculum [87].	2018	RAPIDMINER
<b>Aporte</b>	Este documento estudiará diferentes enfoques que son utilizados por diferentes instituciones de educación superior para integrar los conceptos		



	de minería de datos en su plan de estudios. Demostrando que en varios de esos trabajos se ha hecho uso de weka.		
TR15	Feature-based Restaurant Customer Reviews Process Model using Data Mining [88].	2018	WEKA
<b>Aporte</b>	El propósito de este trabajo de investigación es desarrollar un modelo de software basado en características para analizar las reseñas de los clientes de una organización usando su página de Facebook y proporcionar información valiosa para la toma de decisiones, el desarrollo de la calidad del producto y las mejoras de los procesos. Permitiendo así actividades de ingeniería concurrentes y mejorando la colaboración entre varios departamentos dentro de la organización. Como muestra de un caso de estudio, hemos analizado las opiniones de los clientes de un restaurante utilizando el algoritmo de clasificación J48 y el algoritmo de agrupación K-means para identificar áreas que necesitan mejoras.		
TR16	Comparative Study of Heart Disease Diagnosis Using Top Ten Data Mining Classification Algorithms [89].	2019	WEKA
<b>Aporte</b>	En este estudio, los datos de los registros médicos se recopilan del Hospital Harapan Kita y se utilizan como una muestra de conjunto de datos en esta investigación. Los diez mejores algoritmos de clasificación de minería de datos se utilizan para diagnosticar enfermedades cardíacas a partir de los datos del Hospital Harapan Kita y para examinar su rendimiento al verificar la precisión y la velocidad, trabajando con weka.		
TR17	Data Mining Using Clustering Algorithm as Tool for Poverty Analysis [90].	2019	RAPIDMINER
<b>Aporte</b>	En este estudio, se evaluaron diferentes algoritmos de agrupamiento utilizando el índice de silueta para obtener el mejor algoritmo de agrupamiento para agrupar los hogares y analizar los datos de pobreza. El algoritmo k-means donde $k = 3$ superó a DBSCAN y k-medoids con una silueta de 0.308. El algoritmo produjo tres grupos o grupos y se etiquetó como no pobre, casi pobre y pobre. El resultado puede ayudar a los formuladores de políticas a formular e implementar políticas y programas de reducción de la pobreza que sean claros, razonables, realistas y ejecutables.		
TR18	Research on Customer Deposit Order Based on Data Mining Technology [91].	2019	PYTHON
<b>Aporte</b>	los datos de las instituciones bancarias se analizan y limpian para convertir los datos categóricos en datos numéricos. Luego, para "dividir" el preprocesamiento de los datos después del procesamiento, dividido en conjunto de entrenamiento y conjunto de prueba. En el conjunto de entrenamiento, se establecen LR, Naive Bayes, K-NN, Árbol de decisión, Bosques aleatorios, Extra-árboles, AdaBoost, GBDT, XGBoost, LightGBM y CatBoost (a través del método de búsqueda de cuadrícula GridSearchCV) para estudiar la orden de depósito del cliente tendencia respectivamente.		

	Finalmente, el algoritmo de fusión de apilamiento en Ensemble Learning (el meta-alumno adopta el algoritmo LR) se utiliza para fusionar cada modelo individual para construir el mejor modelo de clasificación de pedidos de depósito a plazo del cliente bancario. Los resultados de la investigación muestran que, usando Python, basado en las diferentes dimensiones de la información del cliente del banco, el algoritmo de fusión de apilamiento basado en el algoritmo Ensemble tiene un mejor efecto de predicción que el modelo único, y el algoritmo de fusión de apilamiento tiene una mejor solidez.		
TR19	Feature-based Facebook reviews process model for e-management using data mining [92].	2019	WEKA
<b>Aporte</b>	Se trata de un sistema que identifica las expresiones de opinión como frases que contienen palabras de opinión, características de opinión y también modificadores de opinión. Estas expresiones se clasifican como positivas, negativas o neutrales. Las expresiones de opinión se identifican y categorizan utilizando técnicas lingüísticas localizadas. Las opiniones se pueden congregan en cualquier nivel de especificidad deseado, es decir, nivel de característica o nivel de producto, nivel de usuario o nivel de servicio, etc.		
TR20	Predicting academic performance of university students from multi-sources data in blended learning [93].	2019	WEKA
<b>Aporte</b>	Se propone predecir el rendimiento académico de los estudiantes universitarios a partir de datos de múltiples fuentes en entornos de aprendizaje multimodales y mixtos utilizando la fusión de datos y la minería de datos. Hemos recopilado datos de 65 estudiantes universitarios y diferentes variables de cuatro fuentes diferentes. En primer lugar, aplicamos fusión de datos y pre procesamiento para crear un conjunto de datos de resumen en formato numérico y categórico. Luego, hemos aplicado diferentes algoritmos de clasificación de cuadros blancos proporcionados por la herramienta de minería de datos de Weka para seleccionar el mejor algoritmo. Finalmente, mostramos el mejor modelo de predicción para ayudar al instructor a tomar medidas correctivas con los estudiantes en riesgo de abandono o fracaso.		
TR21	Research on the Effect of Blended Learning Mode Based on Text Data Analysis [94].	2019	WEKA
<b>Aporte</b>	Este estudio diseña un modo de aprendizaje combinado con cuatro patrones de comunicación y un ciclo de enseñanza efectivo como ideas principales. Utiliza principalmente tecnología de minería de datos para analizar los registros de comunicación del grupo en línea, y se concluye que los estudiantes en la clase piloto en el modo de Aprendizaje combinado tienen una mejor iniciativa de aprendizaje. A través de la comparación de la calificación de la tarea y los resultados del examen final con las clases de		



	control, se demuestran aún más las ventajas del modo de aprendizaje combinado.		
TR22	Fuzzy decision tree for breast cancer prediction [95].	2019	WEKA
<b>Aporte</b>	<p>En este documento, la minería de datos se usó para predecir la etapa del cáncer de seno utilizando un híbrido de lógica difusa y árbol de decisión. La selección de características se utilizó para determinar el mejor atributo en el conjunto de datos de Epidemiología de vigilancia y Resultados finales (SEER). El conjunto de datos consiste en la incidencia de 1975 a 2016, pero el estudio limita el análisis de 2010 a 2016. Se realizan diferentes limpiezas y pre procesamientos de datos. Después de un pre procesamiento exhaustivo de los datos, Se seleccionan seis (6) atributos y una (1) clase objetivo. La comparación del rendimiento muestra que el árbol de decisión difuso logró una mayor precisión del 99,96%, sensibilidad del 99,26% y especificidad del 99,98% que la técnica de clasificación del árbol de decisión. El resultado de la simulación muestra una instancia clasificada correctamente de 165,124, que equivale al 99.97% y solo 351 instancias clasificadas incorrectas o 0.21%. Por lo tanto, un árbol de decisión difuso es más robusto que el clasificador tradicional del árbol de decisión para predecir la etapa del cáncer de mama. 97% y solo 351 instancias clasificadas incorrectas o 0.21%. Por lo tanto, un árbol de decisión difuso es más robusto que el clasificador tradicional del árbol de decisión para predecir la etapa del cáncer de mama. 97% y solo 351 instancias clasificadas incorrectas o 0.21%. Por lo tanto, un árbol de decisión difuso es más robusto que el clasificador tradicional del árbol de decisión para predecir la etapa del cáncer de mama.</p>		
TR23	Machine Learning Based Knowledge Organization in Online Communities [96].	2019	WEKA
<b>Aporte</b>	<p>En este trabajo, se muestra el interés en la clasificación de las publicaciones publicadas por los usuarios en un soporte comunitario en línea sobre la lactancia materna. Esto facilitará el acceso a experiencias compartidas por parte de las usuarias que buscan apoyo (madres que amamantan) y también por parte de profesionales de la salud. Para esta investigación se hizo uso de la herramienta weka.</p>		
TR24	Deep Learning Enabled Predicting Modeling of Mortality of Diabetes Mellitus Patients [97].	2019	PYTHON
<b>Aporte</b>	<p>En este documento, el conjunto de datos MIMIC-III se utiliza para desarrollar y entrenar múltiples modelos que tienen como objetivo predecir la mortalidad de los pacientes con DM. Nuestro modelo de aprendizaje profundo de la red neuronal convolucional produjo un puntaje de 0.885 AUC, sobre todo los modelos de línea de base que construimos, que incluyen árboles de decisión, bosques aleatorios y redes neuronales completamente conectadas. Las entradas para cada modelo se compusieron de tipo de admisión, edad, puntaje de comorbilidad de Elixhauser, mediciones de glucosa en sangre, y rango de glucosa en</p>		

	sangre. Los resultados obtenidos de estos modelos son valiosos para los médicos, pacientes y compañías de seguros.
--	--

Dentro de la metodología a utilizar para el proceso de desarrollo de Minería de Datos Educativa, se encontraron los siguientes resultados descritos en la Tabla XII:

Tabla XI. Trabajos relacionados con la metodología a utilizar

<b>Trabajos Relacionados en base a las metodologías más utilizadas para la MD</b>			
<b>Código</b>	<b>Título</b>	<b>Año de publicación</b>	<b>Metodología utilizada</b>
TR01	Adapting CRISP-DM for idea mining. A data mining process for generating ideas using a textual dataset [98]	2020	CRISP-DM
<b>Aporte</b>	El propósito de este documento es proponer un modelo reutilizable para generar ideas, CRISP-DM, para Idea Mining (CRISP-IM). El diseño y desarrollo de la CRISP-IM se realizan siguiendo el enfoque de la ciencia del diseño. Los CRISP-IM facilita la generación de ideas, mediante el uso de modelado dinámico de temas (DTM), máquina sin supervisión aprendizaje y posterior análisis estadístico en un conjunto de datos de artículos escolares. El CRISP-IM adaptado se puede utilizar para guiar el proceso de identificación de tendencias utilizando literatura académica conjuntos de datos o patentes organizadas temporalmente o cualquier otro texto conjunto de datos de cualquier dominio para generar ideas.		
TR02	Estilos de aprendizaje y minería de datos: un estudio preliminar en el contexto Universitario [99].	2018	CRISP-DM
<b>Aporte</b>	Este artículo presenta hallazgos preliminares de la investigación en curso sobre estilos de aprendizaje y rendimiento académico en estudiantes de ingeniería de la Universidad Pontificia Bolivariana; para ello se ha hecho uso de la minería de datos como técnica para el ordenamiento y análisis de datos. La investigación se articula a la apuesta de la Universidad por prevenir la deserción estudiantil y mejorar las practicas pedagógicas de los docentes universitarios, teniendo como marco de referencia la apuesta del Ministerio de Educación Nacional.		
TR03	An effective research on data mining techniques for intrusion detection & learning classes [100].	2019	CRISP-DM
<b>Aporte</b>	Se propone un ajuste ejecutado utilizando la máquina Hadoop y Map Lessen. Se aumenta el plan arquetípico CRISP-DM en los medios que lo acompañan. Para activarlo, se ensambla un marco real para atornillar la adición del solicitante como un golpe utilizando el abismo Apache y abundancia en el marco de certificado transmitido de Hadoop (HDFS) y posteriormente que administran nuestras metodologías.		

TR04	Comparison of K-Nearest Neighbour and support vector machine for choosing senior high school [101].	2019	KDD
<b>Aporte</b>	<p>El objetivo de esta investigación es comparar el algoritmo K-Nearest Neighbour (KNN) y Support Vector Machine (SVM) que se utilizan para elegir la recomendación de la escuela secundaria superior. La metodología que se utilizó en esta investigación es la minería de datos a través de etapas de descubrimiento de conocimiento en bases de datos (KDD), que consisten en limpieza, integración de datos, selección de datos, transformación de datos, minería de datos, evaluación de patrones y presentación de conocimientos. KNN y SVM son los algoritmos más comunes utilizados en la minería de datos y el sistema de soporte de decisiones. Ya sea KNN o SVM en esta investigación utilizada para clasificar el tipo de escuela secundaria superior por parámetros de entrada, entre otros puntajes de examen nacional, interés de los estudiantes y sugerencia del consejero. Basado en el experimento con varios datos de entrenamiento y pruebas, el resultado muestra que SVM es mejor que KNN. SVM tiene un valor de precisión de alrededor del 97.1%, mientras que KNN tiene un valor de precisión de alrededor del 88.5%. Y también, SVM tiene un tiempo de procesamiento más rápido que KNN.</p>		
TR05	Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia [102].	2019	KDD
<b>Aporte</b>	<p>Se presentan y analizan los resultados de la prueba de la calidad de la educación superior en Colombia – Saber Pro. Se usó la metodología de extracción de conocimiento en bases de datos KDD sobre la cual se construyó una base de datos del desempeño académico del estudiante en áreas asociadas con los contenidos de la prueba Saber Pro, y se utilizó redes neuronales como técnica para la minería de datos. Las redes neuronales permitieron la predicción de los resultados de la prueba Saber Pro con alta exactitud tanto en rangos cualitativos como cuantitativos. Además, se comprobó una correlación entre el desempeño académico y los resultados de Saber Pro. Los hallazgos sugieren que la metodología usada es una excelente guía para el descubrimiento de patrones ocultos en los datos, y permite establecer estrategias de mejora de los resultados de las pruebas Saber Pro que involucren el desempeño académico del estudiante.</p>		
TR06	Identification of factors associated with academic performance in mathematics in the saber 11th tests applying educational data mining [103].	2019	CRISP-DM
<b>Aporte</b>	<p>Este artículo presenta uno de los resultados obtenidos en el proyecto de investigación que tuvo como objetivo aplicar la minería de datos educativos para descubrir los factores asociados con el rendimiento académico de los estudiantes colombianos de secundaria que presentaron la 11 prueba Saber entre los años 2015 y 2016. La información socioeconómica, académica e institucional de esos estudiantes fue seleccionada de las</p>		

	bases de datos ICFES. CRISP-DM se utilizó como metodología. Se construyó, limpió y transformó un repositorio de datos para la minería de datos. Los patrones asociados con el rendimiento académico bueno o malo de los estudiantes se descubrieron en la prueba de matemáticas, utilizando un modelo de clasificación basado en árboles de decisión.		
TR07	Databased learning of influencing factors in order specific transition times [104].	2019	CRISP-DM
<b>Aporte</b>	Este artículo presenta una metodología para identificar factores influyentes basados en datos en un TT específico de orden. Científicamente, La metodología se basa en el Proceso Estándar de la Industria para la Minería de Datos (CRISP-DM) e incluye filtros y envoltorios como métodos de selección de características para la minería de datos (DM). La metodología considera características específicas de la fábrica, identifica características relevantes y no redundantes como factores clave e integra continuamente a los empleados de la fábrica en el proceso de aprendizaje.		
TR08	Determinating student interactions in a virtual learning environment using data mining [105].	2019	CRISP-DM
<b>Aporte</b>	Este artículo se enfoca en determinar las interacciones de los estudiantes en el Curso Virtual de inglés con el Modelo de Educación a Distancia (DEM) en la Universidad de Mumbai, en India. Para este propósito, se realizó un análisis en la base de datos de los estudiantes durante el período académico 2015-2018 para seleccionar los atributos necesarios que permitieron generar un modelo de minería de datos. Posteriormente se realizó un análisis de los métodos de minería comparando cada uno de ellos con el fin de seleccionar el que ayude al desarrollo del proyecto, eligiendo el método Crisp-dm ya que contiene múltiples fases que indican cada actividad a completar, convirtiéndose así en una práctica guía.		
TR09	Análisis del desempeño del profesorado universitario en el uso de MOODLE a través de técnicas de minería de datos: propuestas de necesidades formativas [106].	2018	KDD
<b>Aporte</b>	Este trabajo está orientado al análisis del desempeño de 484 profesores de la Universidad Autónoma de Yucatán (México) durante el período enero - julio del año 2016, que utilizaron la plataforma MOODLE como apoyo tecnológico a su actividad docente, con el objetivo de identificar sus patrones de comportamiento en el uso de las actividades y recursos que contiene esta herramienta tecnológica. La actividad de los profesores en la plataforma "UADY Virtual" fue analizada a través del Método de Extracción de Conocimiento en Bases de Datos (KDD), generando grupos mutuamente excluyentes, reglas de asociación y clasificación que establecen perfiles de profesores y herramientas que deben ser promovidas en esquemas de capacitación. Adicionalmente, se realizó un análisis estadístico del uso de la plataforma. Se determinó que no existen diferencias significativas en el uso de la plataforma MOODLE y el área de conocimiento del profesor, lo		

	que significa que la manera en la que interactúan los profesores con “UADY Virtual” es igual para todas las áreas de conocimiento.		
TR10	The development of a performance appraisal system using Decision Tree analysis and Fuzzy Logic [107].	2018	CRISP-DM, XP
<b>Aporte</b>	Este estudio presenta el desarrollo de un sistema de evaluación del desempeño que tiene como objetivo estudiar los recursos humanos específicos del entorno educativo y resalta el papel de la minería de datos para lograr un desarrollo mejorado de calidad en su Facultad. Los investigadores utilizaron las metodologías CRISP-DM y Extreme Programming, centrándose en generar modelos para el algoritmo del Árbol de decisión, combinado con Fuzzy Logic Controller para predecir el rendimiento del profesorado. Las reglas IF-THEN generadas por J48 se utilizan junto con FLC para predecir el rendimiento del profesorado individual o institucional. Además, la salida generada del prototipo cumple con estándares sustanciales.		
TR11	Parking garage management using classification techniques, fuzzy logic and evolutionary algorithms [108].	2018	KDD
<b>Aporte</b>	Este artículo presenta el análisis de los datos de un garaje para vehículos livianos y pesados para extraer suficiente información que nos permita administrar el estacionamiento y poder predecir la cantidad de espacios libres en un día determinado, así como para encontrar los clientes potenciales. Esto se hizo usando la metodología KDD, para la minería de datos se utilizó la herramienta WEKA, KEEL y Matlab que nos proporcionan diferentes algoritmos como J48, Bayes BayesNet, KNN-C, DT_GA-C y GANN-C para hacer la evaluación. Como resultado, se clasificaron los posibles clientes potenciales de la empresa y, además, se obtiene el mejor resultado con el algoritmo J48, que obtuvo una precisión del 90.43% y un porcentaje de error de 0.4553 utilizando el algoritmo GANN-C.		
TR12	Integrating Data Science into a General Education Information Technology Course: An Approach to Developing Data Savvy Undergraduates [109]	2019	CRISP-DM
<b>Aporte</b>	Muestra un curso como resultado, un curso de TI de encuesta puede proporcionar una educación introductoria integral en ciencias de datos al agregar un módulo de ciencia de datos enfocado en modelado y evaluación, dos pasos clave en el proceso de ciencia de datos. El módulo debe usar software de ciencia de datos para la aplicación, evitando las complejidades de la programación y las matemáticas avanzadas, al tiempo que permite un énfasis en la comprensión conceptual. Como resultado los estudiantes aprendieron el proceso de ciencia de datos CRISP-DM y lo aplicó individualmente problemas y un proyecto grupal integral.		
TR13	Understanding E-learners' Behaviour Using Data Mining Techniques	2019	CRISP-DM
<b>Aporte</b>	Este estudio aborda el comportamiento de los alumnos electrónicos mediante el aprendizaje automático, en particular las reglas de asociación		

	<p>y los clasificadores. Se examinó diferentes técnicas de minería de datos, incluidos los bosques aleatorios, las regresiones logísticas y las redes bayesianas como clasificadores utilizados para predecir las clases de los alumnos electrónicos (alta, media y baja). La novedad de este estudio es que explora y compara el rendimiento de los clasificadores sobre el comportamiento de los alumnos en línea en cuatro variables: levantar la mano, visitar recursos de TI, ver anuncios y el impacto de la discusión en los alumnos electrónicos. Los resultados de este estudio indican un nivel de precisión del 80% obtenido por Bayesian Networks; por el contrario, los bosques aleatorios tienen solo un 63% de nivel de precisión y regresiones logísticas del 58%.</p>		
TR14	A Model for Predicting Disapproval of Apprentices in Distance Education Using Decision Tree [110].	2019	CRISP-DM
<b>Aporte</b>	<p>Este documento propone el MD-PREAD, un modelo que utiliza la técnica del árbol de decisiones para predecir aprendices con riesgo de fracaso. La capacidad de elegir el árbol de decisión como una forma de generar un conjunto mayor para los educadores es lo más destacado de este proyecto. Para el desarrollo se utilizó la metodología crisp-dm, mientras que, para evaluar el modelo, los índices de los clasificadores se compararon y se destacó el algoritmo J48 con un predominio de precisión del 84.5%, precisión del 85.52%. Era concluyó que el modelo MD-PREAD puede ayudar en el pronóstico de grupos en riesgo de fracaso.</p>		
TR15	Embedding naïve Bayes algorithm data model in predicting student graduation	2019	KDD-CRISP-DM
<b>Aporte</b>	<p>El estudio se centra en la aplicación del algoritmo Naïve Bayes en la predicción de la graduación de los estudiantes mediante la generación de un modelo que podría pronosticar e identificar a los estudiantes propensos a no graduarse a tiempo, por lo que las instituciones pueden formular e implementar políticas adecuadas de remediación y retención.</p>		
TR16	A data modeling approach for classification problems: application to bank telemarketing prediction [111].	2019	CRISP-DM
<b>Aporte</b>	<p>En este documento, se presenta un nuevo enfoque de modelado de datos para cinco algoritmos de clasificación comunes para optimizar la predicción de las llamadas objetivo de telemarketing para vender depósitos bancarios a largo plazo. Un banco minorista portugués abordó, desde 2008 hasta 2013, datos sobre sus clientes, productos y atributos socioeconómicos, incluidos los efectos de la crisis financiera. Se ha explorado un conjunto original de 150 características y se retienen 21 características para el enfoque propuesto, incluida la etiqueta. Este documento presenta un nuevo enfoque de modelado que pre procesó por separado cada tipo de características y las normalizó para optimizar el rendimiento de la predicción. Para el desarrollo del trabajo se hace uso de la metodología Crisp-dm, para evaluar el enfoque propuesto, este documento compara los resultados obtenidos con las cinco técnicas de aprendizaje automático más conocidas: Naïve Bayes (NB), Regresión logística (LR), Árboles de decisión</p>		

	(DT), Red neuronal artificial (ANN) y Máquinas de vectores de soporte (SVM) ) y arrojó mejores prestaciones mejoradas para todos estos algoritmos en términos de precisión y medida.		
TR17	Comparison Of Classification Algorithms For The Detection Of Bone Weakness In Students Using Anthropometric Data [112].	2019	CRISP-DM
<b>Aporte</b>	Esta investigación compara cinco algoritmos de minería de datos para predecir la debilidad ósea en estudiantes entre 5 y 18 años de edad. La metodología utilizada para el procesamiento de datos es CRISP-DM. Se discute la precisión de los algoritmos aplicados en los trabajos referenciados con los resultados obtenidos con la herramienta de minería de datos WEKA. Después de hacer la comparación, se determinó que el algoritmo JRip era más preciso.		
TR18	Optimization Model of Physical Education Teaching Evaluation Based on Association Rule Algorithm [113].	2019	KDD
<b>Aporte</b>	Este trabajo se centra en la investigación del algoritmo de reglas de asociación y su aplicación. En un mejor entorno de hardware y software y con la ayuda de algunos datos en la base de datos de transacciones de supermercados reales, el documento adopta Microsoft SQL Server 2005 y VB. NET se utiliza como plataforma de desarrollo para construir un sistema simple de minería de reglas de asociación y aplicar el algoritmo BMSL a la generación de reglas de asociación, así como también la metodología KDD.		
TR19	Predicting master's applicants performance using KDD techniques [114].	2017	KDD
<b>Aporte</b>	Este documento presenta un análisis de las características académicas de los solicitantes de educación de maestría. Se utilizó para identificar un conjunto de características que podrían ayudar a predecir la producción científica de un solicitante. A través del proceso de descubrimiento de conocimiento en la base de datos, Se obtuvo un conjunto de reglas de correlación entre las características y el desempeño esperado del solicitante. La recopilación de datos se realizó en la plataforma Lattes (una plataforma brasileña para currículos científicos). Luego, después del pre procesado, se optimizó utilizando la partición alrededor de los medoides con la distancia de Gower. Los experimentos se realizaron entrenando siete algoritmos de clasificación en dos enfoques. El primero fue un enfoque ingenuo, haciendo que el análisis sea lo más simple posible y el segundo, desarrollar un proceso optimizado de DM. La optimización utilizó el proceso de agrupación para crear una nueva clase principal para los clasificadores. Los resultados demuestran que el enfoque optimizado fue capaz de lograr una precisión de alrededor del 85% utilizando el Bosque aleatorio, mientras que el ingenuo tiene aproximadamente el 44%. A través del algoritmo PAM, fue posible determinar los mejores atributos y resaltar las reglas principales. La conclusión es que la categoría de Universidad y su producción previa		

	fueron los atributos más relevantes. Muestra que es posible predecir la productividad de los estudiantes de maestría.		
TR20	Knowledge Discovery in Data Bases: a Case Study in a Private Institution of Higher Education [115].	2018	KDD
<b>Aporte</b>	Este documento tiene como objetivo presentar una metodología para descubrir el conocimiento en bases de datos (Knowledge Discovery in Databases; KDD) que se puede utilizar en varias áreas para clasificar nuevas instancias. Su aplicación se muestra en cursos de una institución privada de educación superior para verificar la satisfacción de los estudiantes en relación con la calidad y los servicios ofrecidos. Basado en el proceso KDD, inicialmente se realizó un análisis exploratorio de datos y, después de eso, se aplicó tres técnicas de minería de datos: binario de regresión logística (LRB), Generación de superficie que minimiza los errores a través de un modelo matemático de programación lineal (GSME-PL) y la función lineal discriminadora de Fisher (FDLF). Se analizaron 885 instancias, con 12 variables y una salida (satisfacción). A través de los resultados obtenidos, se puede concluir que los atributos relacionados con los "maestros" son los más importantes, y para el caso abordado, RLB fue la técnica con la tasa de precisión más alta (92.2%).		
TR21	An educational model based on Knowledge Discovery in Databases (KDD) to predict learner's behavior using classification techniques [116].	2016	KDD
<b>Aporte</b>	Este documento examinó la historia de los estudiantes de acceder a los datos del Sistema de Gestión de Aprendizaje (LMS) de la universidad. Las técnicas de clasificación se utilizan para construir un modelo educativo basado en el descubrimiento de conocimiento en bases de datos (KDD) para predecir el comportamiento del alumno. Identificó el influyente más valioso para los resultados de aprendizaje de los alumnos; generó modelos de predicción utilizando el algoritmo del árbol de decisión J48 y la regresión lineal múltiple; y determinó qué tan probable es que los estudiantes de Educación a Distancia (DE) obtengan una calificación de "Aprobado" en un curso determinado que puede ofrecer información vital a los maestros y administradores universitarios para la planificación del programa y las estrategias de apoyo al estudiante. Los proponentes realizaron experimentos para predecir la calificación final de los estudiantes en función de su historial de acceso a los datos en el LMS de la universidad. Basado en el modelo derivado, el puntaje obtenido de la participación en las actividades en línea fue el factor de influencia más valioso para los resultados de aprendizaje de los estudiantes de.		
TR22	BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset [117].	2020	KDD
<b>Aporte</b>	Para resolver los problemas de baja precisión e ingeniería de características en la detección de intrusos, se propone un modelo de detección de anomalías de tráfico BAT. El modelo BAT combina BLSTM		



(memoria bidireccional a corto y largo plazo) y mecanismo de atención. El mecanismo de atención se utiliza para filtrar el vector de flujo de red compuesto por vectores de paquetes generados por el modelo BLSTM, que puede obtener las características clave para la clasificación del tráfico de red. Además, adoptamos múltiples capas convolucionales para capturar las características locales de los datos de tráfico. Como se usan múltiples capas convolucionales para procesar muestras de datos, nos referimos al modelo BAT como BAT-MC. El clasificador softmax se usa para la clasificación del tráfico de red. El modelo de extremo a extremo propuesto no utiliza ninguna habilidad de ingeniería de características y puede aprender automáticamente las características clave de la jerarquía. Puede describir bien el comportamiento del tráfico de la red y mejorar la capacidad de detección de anomalías de manera efectiva.
---

## **2.9. Análisis de Resultados**

En esta sección se realizó un análisis de los 67 trabajos seleccionados que dan respuestas a las preguntas de investigación, obtenidos a través de todo el proceso de RSL, incluyendo los criterios de inclusión y exclusión para la misma.

### **2.9.1. Presentación de Resultados**

Dentro de los resultados obtenidos se logró identificar que los factores que influyen en la deserción estudiantil universitaria son los Académicos, Sociales y Económicos, la mayoría de estos se argumentan en que los aspectos educativos tienen mucho que ver para “desanimar” a los estudiantes sobre la Carrera que escogió, esto a futuro hace que el alumno deserte sus estudios, puesto que se considera que en Carreras de ingeniería la enseñanza se muestra más robusta ya que su relación con asignaturas relacionadas con las matemáticas es muy estrecha en comparación con otras Carreras.

Dentro de estos trabajos relacionados se ha determinado una lista de algoritmos de Clasificación como los que se muestran en la Figura 26, donde se identifica que el algoritmo más utilizado para este tipo de investigación es el Árbol de decisión (Decision tree) con el 35% del total.

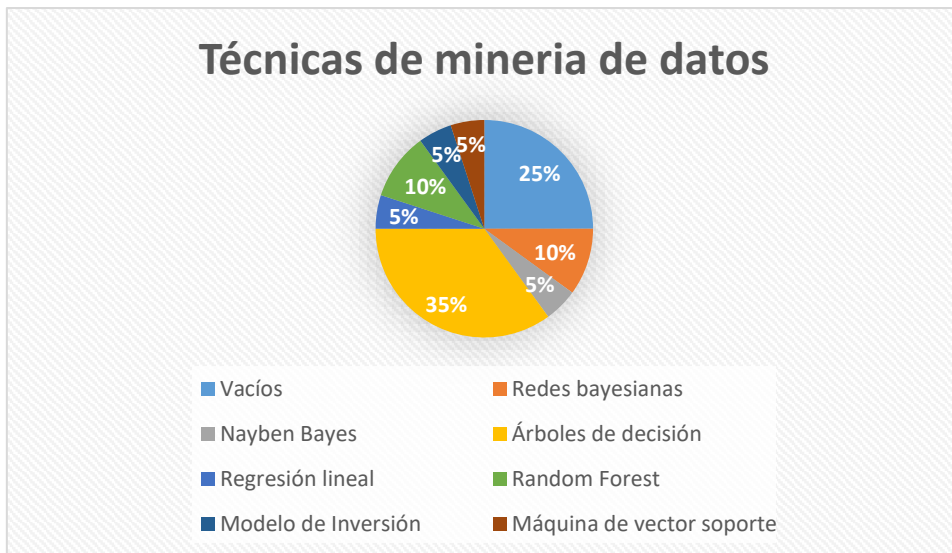


Figura 26. Técnicas de minería de datos

En la Figura 27 muestra las herramientas a utilizar para la ejecución de la minería de datos, estas son: Weka, RapidMiner y Python, en ella se logra identificar que la herramienta que más se utiliza es Weka con un 76% equivalente a 19 del total de trabajos relacionados acorde a las herramientas.

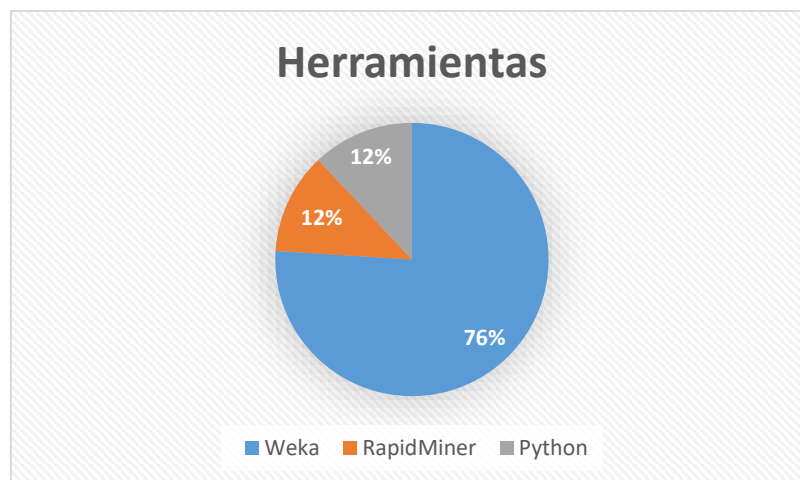


Figura 27. Herramientas para la ejecución de MD.

En la Figura 28 se muestran las metodologías que más se utilizan para el desarrollo de Minería de datos; dentro de los trabajos obtenidos en RSL en relación a la metodología se puede identificar como la más implementada a CRISP-DM con un 59% lo que corresponde a 13 trabajos relacionados del total.

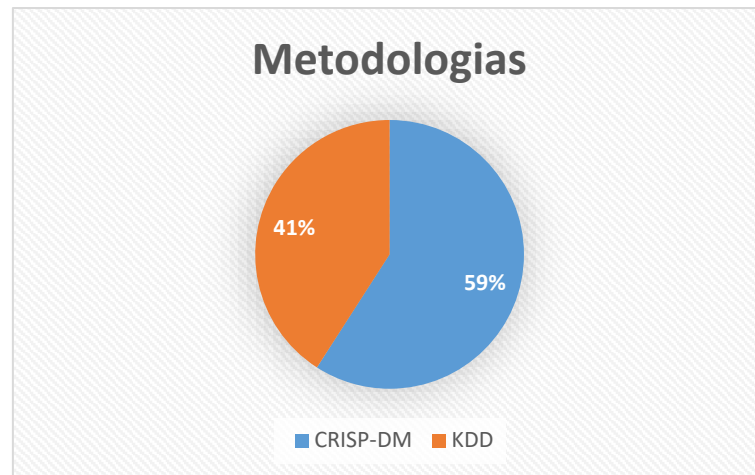


Figura 28. Metodologías para el proceso de desarrollo de Minería de Datos.

### 3. Conclusiones

En base a los resultados obtenidos con la información relevante, se pueden presentar las siguientes conclusiones:

- Identificar los factores que influyen en el abandono de los estudiantes universitarios dentro de los registros académicos de los mismos, permite a las autoridades correspondientes conocer la situación de la institución y en base a ellos tomar decisiones que ayuden a mitigar la deserción universitaria.
- Para obtener resultados favorables es necesario aplicar varias técnicas de minería de datos esto permitirá tener respuestas acercadas a la realidad.
- Utilizar la metodología adecuada de acuerdo a las necesidades requeridas para el desarrollo de minería de datos.

## Anexo 2. Análisis Exploratorio de los datos

El presente documento muestra el análisis y la exploración de los datos, esto consiste en realizar un análisis estadístico gráfico y numérico de las variables del tema de estudio con el fin de tener una idea inicial de la información contenida en el conjunto de datos, así como detectar la existencia de posibles errores en la codificación de estos.

Para ello se ha hecho uso de la del lenguaje Python, el cual nos permite de acuerdo con las necesidades requeridas elaborar un análisis de cada una de las variables presentes dentro del dataset. El cual cuenta con un total de 36 363 registros con 51 variables, tal y como se muestra en la Figura 29.

```

↳ El archivo tiene el siguiente número de filas y columnas: (36362, 51)

Las columnas son: Index(['facultad', 'carrera', 'periodo_lectivo', 'nivel', 'modalidad',
'cedula', 'apellidos', 'nombres', 'fecha_nacimiento', 'genero',
'estado_civil', 'etnia', 'sector_procedencia', 'nacionalidad',
'pais_nacimiento', 'provincia_nacimiento', 'canton_nacimiento',
'ciudad_nacimiento', 'direccion_nacimiento', 'pais_actual',
'provincia_actual', 'canton_actual', 'parroquia_actual',
'ciudad_actual', 'direccion_actual', 'trabaja', 'ingreso_estudiante',
'numero_hijos', 'colegio', 'tipo_colegio', 'pais_colegio',
'provincia_colegio', 'canton_colegio', 'oferta_academica', 'ciclo',
'numero_matricula', 'malla_curricular', 'estado_matricula', 'paralelo',
'jornada', 'asignatura', 'estado_asignatura', 'obligatoria',
'arrastrable', 'asistencia_obligatoria', 'nota_ingresada',
'nota_ponderada', 'porcentaje_asistencias', 'promedio_matricula',
'homologada', 'observacion_homologacion'],
dtype='object')

```

*Figura 29. Variables totales dentro del conjunto de datos inicial.*

Cada una de las variables se detallan a continuación en la Tabla XIII, considerando el tipo de cada una de estas:

*Tabla XII. Detalle de cada variable presente en el conjunto de datos inicial*

Variable	Descripción	Tipo
facultad	Describe la Facultad a la que pertenecen las Carreras de los estudiantes dentro de la UNL.	Texto
carrera	Describe el nombre de la Carrera a la que pertenece el estudiante.	Texto
periodo_lectivo	Nombre que se le asigna a la fecha del ciclo académico.	Numérico
nivel	Representa el régimen de los estudiantes dentro de la universidad.	Texto



modalidad	Modalidad de estudios de las Carreras	Texto
cedula	Cédula de identidad de los estudiantes	Texto
apellidos	Representa los apellidos de los estudiantes	Texto
nombres	Representa los nombres de los estudiantes	Texto
fecha_nacimiento	Contiene la fecha de nacimiento de los estudiantes	Texto
genero	Genero de los estudiantes (Masculino o Femenino)	Texto
estado_civil	Estado civil de los estudiantes	Texto
Etnia	Esta variable representa la etnia de los estudiantes	Texto
sector_procedencia	Describe el sector de procedencia de los estudiantes	Texto
nacionalidad	Describe el estado al que pertenece el estudiante	Texto
pais_nacimiento	Describe el país de nacimiento de los estudiantes	Texto
provincia_nacimiento	Describe la provincia de nacimiento de los estudiantes	Texto
canton_nacimiento	Describe el cantón de nacimientos de los estudiantes	Texto
ciudad_nacimiento	Describe la ciudad de nacimiento de los estudiantes	Texto
direccion_nacimiento	Describe la dirección de nacimiento de los estudiantes	Texto
pais_actual	Describe el país actual de los estudiantes	Texto
provincia_actual	Describe la provincia actual de los estudiantes	Texto
canton_actual	Describe el cantón actual de los estudiantes	Texto
ciudad_actual	Describe la ciudad actual de los estudiantes	Texto
direccion_actual	Describe la dirección actual de los estudiantes	Texto
trabaja	Describe si los estudiantes trabajan o no.	Texto
ingreso_estudiante	Describe el monto total de ingresos de los estudiantes	Numérico



numero_hijos	Describe el número de hijos por parte de los estudiantes	Numérico
colegio	Describe a la institución de formación de los estudiantes	Texto
tipo_colegio	Describe el tipo de institución de los estudiantes, es decir, público, privado o fisco misional.	Texto
pais_colegio	Describe el país donde se encuentra el colegio de los estudiantes	Texto
provincia_colegio	Describe la provincia de colegio de los estudiantes	Texto
canton_colegio	Describe el cantón de colegio de los estudiantes	Texto
oferta_academica	Describe la oferta académica para el ciclo académico	Texto
ciclo	Describe el ciclo que cursan los estudiantes	Numérico
numero_matricula	Describe el número de matrícula de los estudiantes	Numérico
malla_curricular	Describe la malla a la que pertenecen los estudiantes	Texto
estado_matricula	Describe el estado de matrícula de los estudiantes	Texto
paralelo	Describe el paralelo al que pertenecen los estudiantes	Texto
jornada	Muestra la jornada en la que los estudiantes reciben clases	Texto
asignatura	Describe a cada una de las asignaturas que deben aprobar los estudiantes	Texto
estado_asignatura	Representa el estado de la asignatura de los estudiantes es decir si el alumno aprueba o reprueba.	Texto
obligatoria	Describe el estado obligatoriedad de las asignaturas.	Texto
arrastrable	Describe si la asignatura que cursan los estudiantes es arrastrable o no.	Texto
asistencia_obligatoria	Describe la asistencia obligatoria de las asignaturas por parte de los estudiantes.	Texto

nota_ingresada	Describe la calificación de la asignatura ingresada cursada por los estudiantes.	Numérico
nota_ponderada	Describe el 20% de la nota ingresada por cada asignatura.	Numérico
porcentaje_asistencia	Describe el % de asistencia a cada asignatura por los estudiantes	Numérico
promedio_matricula	Describe el promedio de la matrícula de los estudiantes	Numérico
homologacion	Describe si la asignatura cursada por el estudiante es homologada o no.	Texto
observacion_homologacion	Describe si existe alguna observación dentro del proceso de homologación	Texto

Una vez revisada cada uno de las variables identificando el tipo a la que pertenecen se procede a realizar un análisis estadístico con Python para conocer el estado de las mismas, identificando cuáles serán las adecuadas para el caso de estudio.

La variable **facultad** muestra un solo resultado es decir su único valor a tomar es de FEIRNNR, por lo que no es necesaria dentro del conjunto de datos.

La variable **Carrera** indica dos valores, Carrera de Sistemas y Carrera de computación, en la Figura 30 se muestra que el 91.2% de los estudiantes pertenecen a Sistemas y el 8.8% a la Carrera de computación.

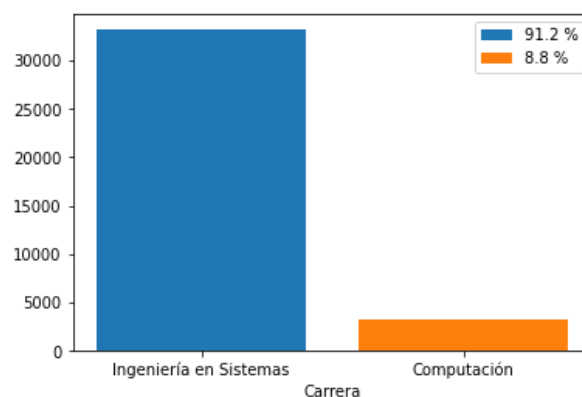


Figura 30. Carreras presentes en el conjunto de datos

La Figura 31 indica a la variable **periodo\_lectivo** esta representa el periodo dado en las clases y se identifican 13 agrupaciones, donde se puede visualizar que existe más consistencia de los datos desde los años 2016 hasta el 2020.

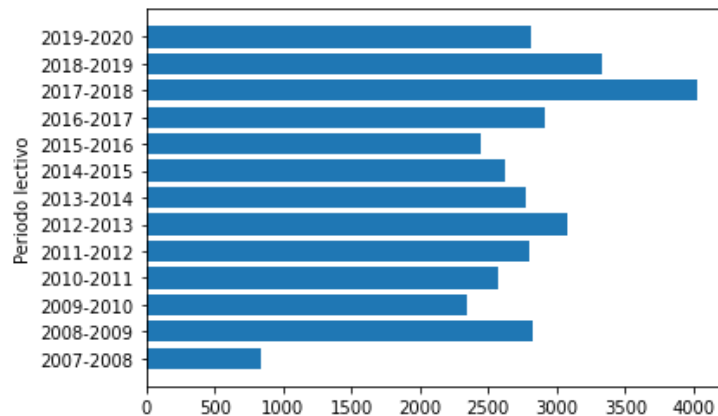


Figura 31. Periodo lectivo dentro del conjunto de datos

La variable **nivel** indica el régimen de las mallas dentro de las Carreras, identificando al 91,2% al régimen 2009 (este vendría a ser la Carrera de Sistemas) y el 8,8% correspondiente al régimen 2013(Carrera de Computación).

Dentro de la variable **modalidad** de estudios dentro de las Carreras se identifica un solo valor que es presencial. De la misma manera la variable **cédula** es el identificador único de cada uno de los estudiantes registrándose en este caso un total de 1800 registros, se debe tomar en cuenta que los datos proporcionado cuenta con un registro por asignatura, es decir si un estudiante ha llegado hasta 5 ciclo y hasta el momento se contabilizan 40 asignaturas, dicho estudiante constará con el número total de asignaturas proporcionadas cuando este debería constar únicamente por un solo ID. Por otro lado, los **apellidos** y **nombres** de los estudiantes son datos que pueden ser reemplazados por la cédula ya que en este caso no es de interés conocer ninguno de estos valores; la **fecha\_nacimiento** es otra variable que no se puede considerar sin embargo proporciona información necesaria para obtener una variable nueva como es la edad.

La Figura 32, indica que dentro de este tipo de Carreras es más tomada por el **género** masculino con un 69.3% mientras que solo un 30.7% pertenecen al género femenino.



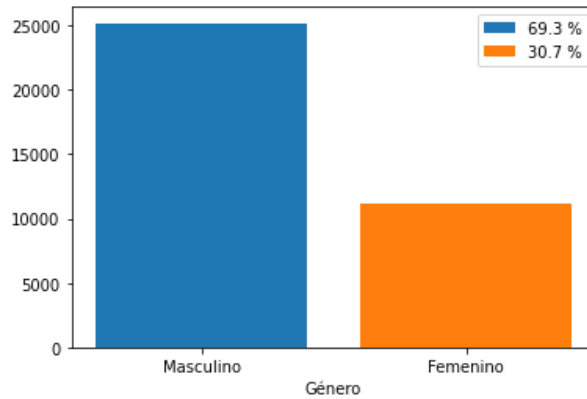


Figura 32. Género de los estudiantes

La variable **estado\_civil** muestra el estado de los estudiantes (Ver Figura 33), en este caso el resultado que predomina es Soltero con un 83.57%, sin dejar a un lado la opción de Casado con un 14.91%, mientras que Divorciado(a), Unión libre y Viudo corresponden 1.44%, 0.05% y 0.03% respectivamente.

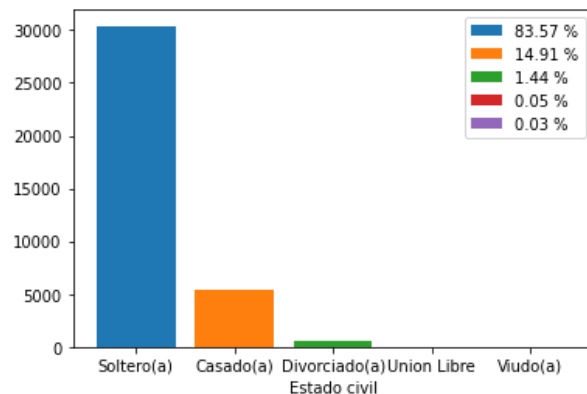


Figura 33. Estado civil de los estudiantes

La Figura 34 muestra a los diferentes valores que **Etnia** puede tomar, como son, Mestizo con un 97.07%, Indígenas con el 1.88%, Blanco con un total de 0.44%, mientras que mulato, montubio y negro abarcan un 0.61%.

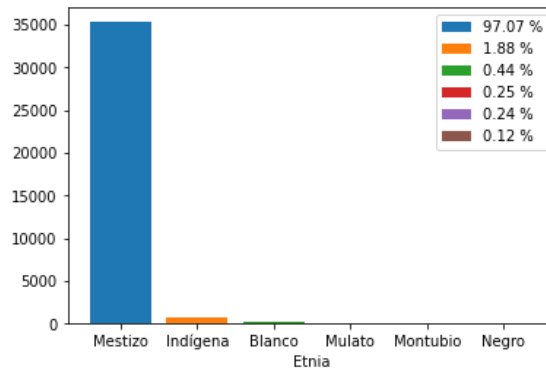


Figura 34. Etnia con la que se identifican los estudiantes.

La variable **sector\_procedencia** hace referencia al sector de donde provienen los estudiantes aquí se identifica que el 92.9% son del sector Urbano mientras que el 7.1% pertenecen al sector Rural (Ver Figura 35).

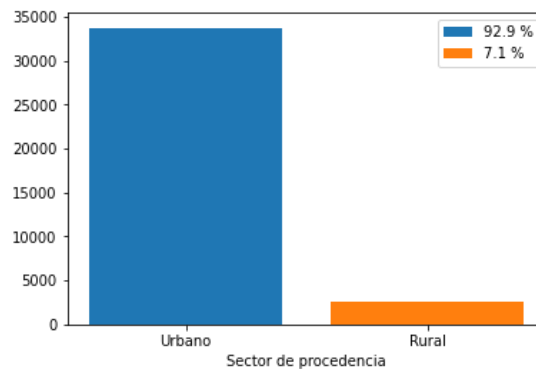


Figura 35. Sector de procedencia de los estudiantes.

La Figura 36 detalla la **nacionalidad** a la que pertenece cada estudiante, en este caso la nacionalidad ecuatoriana es la que sobresale con un 99.884% en comparación con la colombiana y peruana (0.116%).

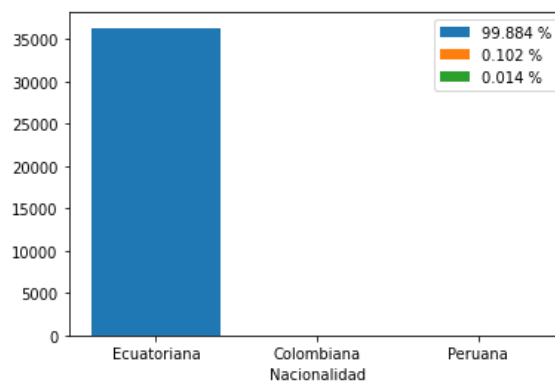


Figura 36. Nacionalidad a la que pertenecen cada estudiante.

Para **pais\_nacimiento** se refleja cada uno de los países donde nacieron cada uno de los estudiantes, dentro de ellos se tiene a Ecuador, Colombia, España y Perú.

La variable **provincia\_nacimiento** demuestra a cada una de las provincias del Ecuador donde son oriundos cada uno de los habitantes, la Figura 37 refleja a Loja como la primera provincia que abarca el mayor número de estudiantes nacidos en esta.

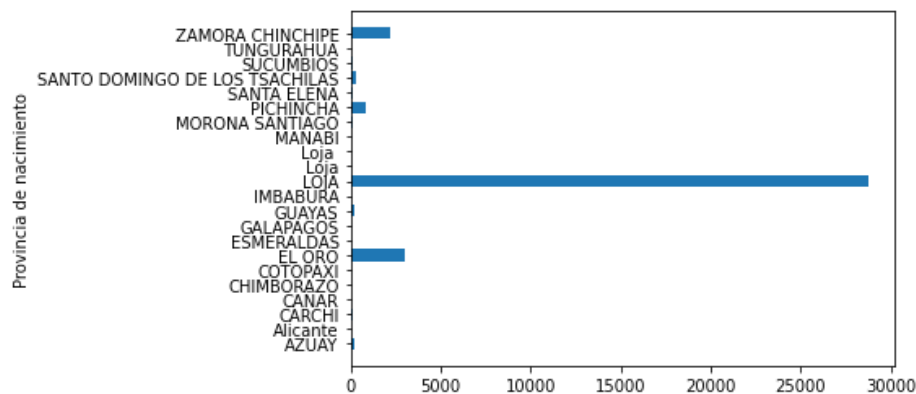


Figura 37. Provincia de nacimiento de los estudiantes

Para las variables **canton\_nacimiento**, **ciudad\_nacimiento**, **direccion\_nacimiento**, **pais\_actual**, **provincia\_actual**, **canton\_actual**, **ciudad\_actual** y **direccion\_actual**, hacen referencia únicamente a donde vivían y viven los estudiantes, estas no indican relevancia de acuerdo al contexto de la investigación.

En la Figura 38, se detalla el porcentaje de los estudiantes que trabajan siendo este en un 91.4% no trabaja mientras que el 8.6% si lo hace.

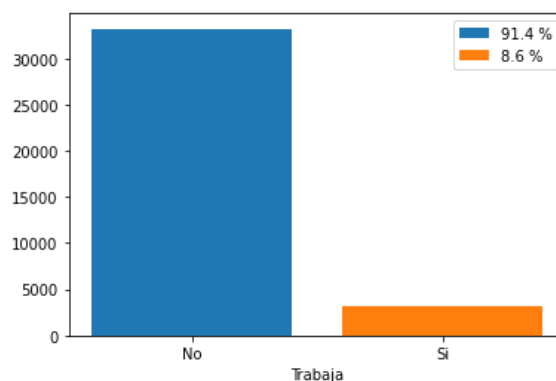


Figura 38. Estado de trabajo de los estudiantes.

La variable **ingreso\_estudiante** no presenta un valor aceptable para el caso de estudio, según la Figura 39 y en relación con la Figura 38 se puede visualizar un bajo ingreso ya que el número de estudiantes que trabajan es mínimo.

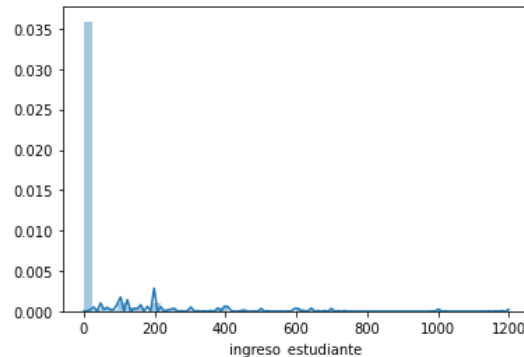


Figura 39. Ingreso mensual de los estudiantes.

La variable **numero\_hijos** muestra un promedio del número de hijos que tienen los estudiantes, considerando desde 0 a 10 hijos, en donde se puede ver que existe un número elevado de personas que no tienen hijos mientras que lo que le sigue es entre 1 y 2 hijos, los valores restantes cuentan con resultados mínimos (Ver Figura 40).

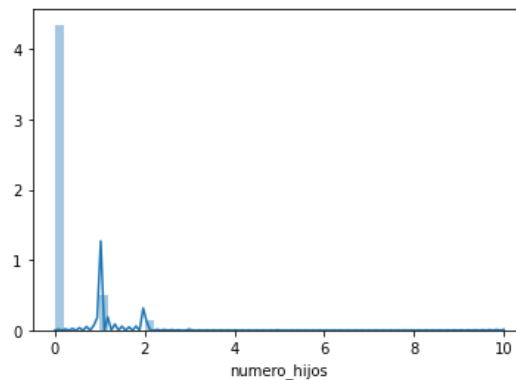


Figura 40. Número de hijos bajo la responsabilidad de los estudiantes

Para el **colegio**, **país\_colegio**, **provincia\_colegio** y **cantón\_colegio** existe un conjunto de registros cualitativos finitos que en este caso no es de interés para la investigación; sin embargo, el **tipo\_colegio** representa una variable importante para identificar si el problema se debe al tipo de la institución en la que los estudiantes se formaron, teniendo los siguientes resultados como lo indica la Figura 41.

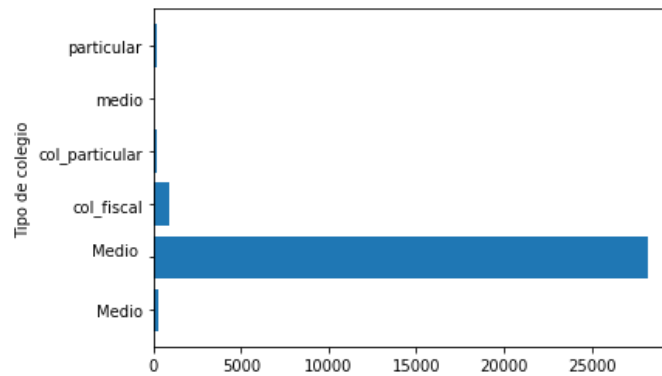


Figura 41. Tipo de colegio en la que se formaron los estudiantes

Dentro de la variable **oferta\_academica** existen registros como Cursos Intensivos marzo 2019 FEIRNNR, Pregrado Marzo – Agosto 2015 y entre otros registros que por la manera en que se llevaban los datos académicos dentro del SGA (Sistema de Gestión Académica), estos no se encuentran correctos e incluso hay valores nulos o incompletos.

La Figura 42 muestra el número de **ciclos** en donde se visualiza un número total de estudiantes que han pasado por los diferentes ciclos (1...10). Dentro de esto se puede manifestar que existe un número elevado que ingresan, sin embargo, a medida que cursan los ciclos el número de estudiantes se reduce a la mitad o en sus peores casos a obtener un número que demuestra una situación preocupante para las autoridades correspondientes. Los últimos ciclos (11-12) hacen referencia a la UTE y Examen Complexivo ciclos asignados con ese nombre para aquellos estudiantes que por motivos diversos no avanzan a realizar si TT dentro del tiempo establecido se los asigna para estas actividades.

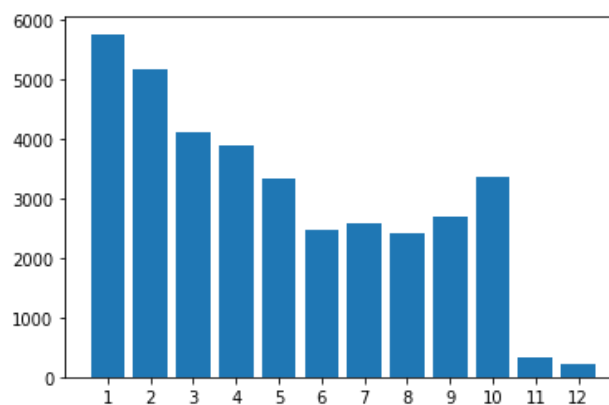


Figura 42. Ciclos por los que los estudiantes deben pasar para culminar su Carrera.

El **numero\_matricula** es asignado a cada estudiante cada que realiza la petición de matrícula para un nuevo ciclo o asignatura en particular, esto genera una cantidad considerable de registros por la cantidad de estudiantes que se ha tenido y tiene.

Dentro de la **malla\_curricular** existen cuatros mallas curriculares específicas (Ver Figura 43), 2014 con un total del 58.75% esta es utilizada para los estudiantes de la Carrera de Computación, la malla 2011 tiene el 17.42% de registros, para esta están los estudiantes que pertenecían al sistemas Modular SAMOT, donde no se llevaba un control automatizado de los datos académicos de cada estudiante, dentro de las mallas Pensum Régimen 2013 y 2013 con 14.33% y 8.84% respectivamente, se encuentran los estudiantes que ya estaban dentro de una nueva actualización del SGA, es decir algunos registros se realizaron de manera automatizada por ende los registros muestran más consistencia y más lógica.

Como se mencionó anteriormente para los estudiantes que no alcanzan a graduarse dentro de los 10 ciclos académicos se les brinda la oportunidad de hacerlo por medio de un Examen Complexivo o por medio de la UTE.

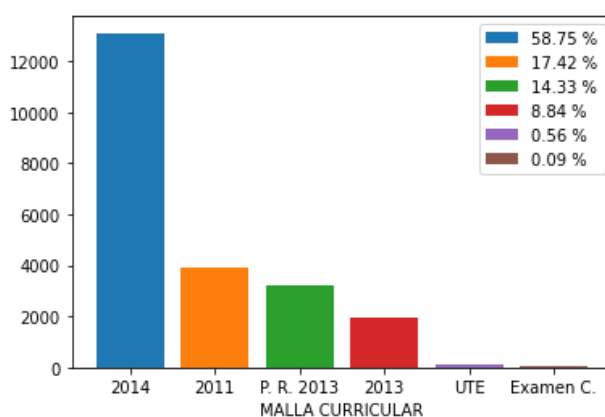


Figura 43. Mallas Curriculares para los estudiantes.

Dentro de la variable **estado\_matricula** se encuentra que únicamente el 69.03% de los estudiantes aprueban las asignaturas, el 24.88% reprueban, las personas que están matriculadas con un 4.56% y el 1.53% se encuentran reubicadas (Ver Figura 44).

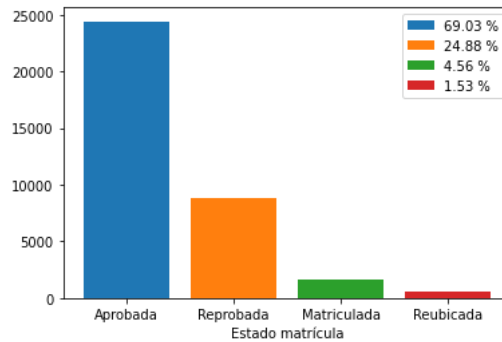


Figura 44. Estado de la matrícula perteneciente a los estudiantes.

La variable **asignatura** muestra un amplio conjunto de valores de los cuales existe mucha inconsistencia debido a que hasta el año 2016 antes que se automatizaran los procesos, los nombres de las asignaturas están con errores de tipeo, por ejemplo “PROGRAMACION” y “PROGRAMACIÓN”, otros con valores vacíos, y dentro de asignaturas existen también cursos o talleres que los hacían constar como asignaturas cuando no era así.

La variable **paralelo**, indica el paralelo asignado a cada uno de los estudiantes, estos pueden ser A, B, y estos esta subdivididos en paralelos para estudiantes con asignaturas arrastrables; **jornada** implica una sola alternativa que en ese caso como la Carrera se da un sola, esta tomaría el valor de matutina. Las variables **obligatoria y arrastrable**, hacen referencia a que si la asignatura que está recibiendo de acuerdo a la malla curricular es obligatoria o no y con lo mismo si esta se puede arrastrar o no. Para las variables **asistencia\_obligatoria, porcentaje\_asistencia, promedio\_matricula, nota\_ingresada, nota\_ponderada**, son las características que acompañan a la asignatura, aquí se muestra calificaciones tanto en asistencia como en promedio general de la misma, esta también presenta la posibilidad de una **homologación**, y si presenta especificar alguna observación presente durante el proceso de homologación dentro de la variable **observacion\_homologacion**.

A nivel del investigador estas últimas variables no aportan ninguna información que permita identificar que en base a ellas un estudiante apruebe a repruebe una asignatura o en cualquier relación que desee encontrar o saber.



## Anexo 3. Reporte de Minería

### 1. Definición del Proyecto

A nivel mundial nos enfrentamos con un alto número de deserción académica, según UNESCO existe un 40% de abandono universitario, todas las instituciones de nivel superior buscan minorar las cifras de deserción, considerando que en Carreras técnicas es donde más existe este inconveniente, así como también el bajo número de mujeres dentro de las mismas. Se busca encontrar toda información (patrones) dentro de los registros académicos de los estudiantes, que permitan conocer cuáles son los factores que influyen al abandono de su carrera. Para ello se hace uso de la Minería de Datos Educativa la cual servirá para descubrir información necesaria que permita mitigar la deserción académica a nivel superior.

### 2. Encuesta Literaria

Para la obtención de la información dentro de los registros académicos se ha realizado uso de los siguientes algoritmos, descritos a continuación:

- **A priori:** Busca ítemsets frecuentes usando generación de postulantes. Su nombre se debe a que usa conocimiento a priori para la generación de ítemsets frecuentes. Este algoritmo se resume en dos pasos importantes:
  - a. Genera todos los ítemsets que contienen un solo elemento, utilizando a estos para generar ítemsets que contengan dos elementos, y así sucesivamente. Se toman todos los posibles pares de ítems que cumplen con las medidas mínimas de soporte inicialmente preestablecidas; esto permite ir eliminando posibles combinaciones: aquellas que no cumplan con los requerimientos de soporte no entrarán en el análisis.
  - b. Generación de las reglas revisando que cumplan con el criterio mínimo de confianza y elevación (lift). Observando que, si una conjunción de consecuentes de una regla cumple con los niveles mínimos de soporte y confianza, sus subconjuntos (consecuentes)



también los cumplen; en el caso contrario, si algún ítem no los cumple no tiene caso considerar sus superconjuntos [44].

- **Simple K-means:** La idea principal es definir  $k$  centroides (uno para cada grupo) y luego tomar cada punto de la base de datos y situarlo en la clase de su centroide más cercano. Después, se recalcula el centroide de cada grupo y se vuelve a distribuir todos los objetos según el centroide más cercano [118].

Cada clúster por tanto es caracterizado por su centro o centroide que se encuentra en el centro o el medio de los elementos que componen el clúster. Kmeans es traducido como K-medias. O un conjunto de objetos  $D_n = (x_1, x_2, \dots, x_n)$ , para todo el  $i, x_i$  reales y  $k, v_1$ , los centros de los  $K$  clúster. El algoritmo del K-means se realiza en 4 etapas [119]:

**Etapas 1:** Elegir aleatoriamente  $K$  objetos que forman así los  $K$  clústers iniciales. Para cada clúster  $k$ , el valor inicial del centro es  $= x_i$ , con los  $x_i$  únicos objetos de  $D_n$  pertenecientes al clúster.

$$s = \operatorname{argmin} \|x_k - x\|^2$$

**Etapas 2:** Reasigna los objetos del clúster. Para cada objeto  $x$ , el prototipo que se le asigna es el que es más próximo al objeto, según una medida de distancia, (habitualmente la medida euclidiana).

**Etapas 3:** Una vez que todos los objetos son colocados, recalcular los centros de  $K$  clúster. (los baricentros).

**Etapas 4:** Repetir las etapas 2 y 3 hasta que no se hagan más reasignaciones. Aunque el algoritmo termina siempre, no se garantiza el obtener la solución óptima. En efecto, el algoritmo es muy sensible a la elección aleatoria de los  $K$  centros iniciales. Esta es la razón por la que, se utiliza el algoritmo del K-means numerosas veces sobre un mismo conjunto de datos para intentar minimizar este efecto, sabiendo que a centros iniciales lo más espaciados posibles dan mejores resultados.

### 3. Métodos

Para llevar a cabo el correcto uso de la Minería de Datos dentro del TT se hizo uso de la metodología CRISP-DM, la misma que se detalla a continuación.

### 3.1. Metodología CRISP-DM

CRISP-DM es la metodología de referencia más utilizada en el desarrollo de proyectos de minería de datos en los ambientes académico [31]. Organiza el desarrollo de un proyecto de Data Mining en una serie de fases o etapas, con tareas generales y específicas que permitan cumplir con los objetivos del proyecto. Estas fases funcionan de manera cíclica e iterativa, pudiendo regresar desde alguna fase a otra anterior [120].

Para el desarrollo del TT se ha determinado el uso de la metodología CRISP-DM, basándose en [32], en el cual se abordaron las siguientes fases:

#### **FASE I. Comprensión del Negocio**

Esta fase se centra en comprender cada uno de los objetivos y los requerimientos necesarios para la elaboración y ejecución del TT desde la perspectiva del negocio, de esta manera permitirá convertir este conocimiento en un problema de minería de datos educativa, así como también en un plan designado a encontrar una solución que permita cumplir con los objetivos.

#### **FASE II. Comprensión de los Datos**

En esta fase se llevó a cabo la recolección de los registros académicos de los estudiantes de la Carrera de Ingeniería en Sistemas/Computación para una mejor comprensión de estos, ya que es la primera etapa para posteriormente realizar un análisis minucioso y con ello identificar como se encuentran estructurados los datos, este proceso se hizo de manera visual en Excel aplicando filtros permitiendo una mejor comprensión de cada uno de ellos.

#### **FASE III. Preparación de los Datos**

Para la ejecución de esta fase se llevó a cabo la preparación de los datos con los cuales se puedan trabajar con las técnicas de minería de datos seleccionada, para ello fue necesaria la estructura de datos de la fase anterior, de la misma manera se llevó a cabo una limpieza de los datos

para poder generar el modelo de minería de datos, tratando de obtener una data sin valores nulos o vacíos.

De la misma manera, con la estructura de datos seleccionada en base a variables tales como fecha\_nacimiento se generaron nuevos atributos a partir de los que ya existían (edad), por otra parte, se realizó agrupación de campos con características comunes para finalmente transformar los datos sin que exista modificaciones en su significado, de esta manera de ajustan a las técnicas de minería de datos arrojando mejores resultados.

#### **FASE IV. Modelado**

En esta fase se escogieron las técnicas de minería de datos más apropiadas para la pregunta de investigación. Una vez escogida la técnica se debe asegurar que se disponga de los datos necesarios y en el formato que la técnica lo necesite, lo cual se ejecutó en la fase anterior.

#### **FASE V. Evaluación**

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis [32].

### **4. Conjunto de Datos**

Para lograr una estructura de datos consistentes que permita aplicar la técnica de mi minería de datos correcta y para conseguir los resultados favores acorde a los objetivos, se realizó primeramente la petición pertinente de los datos al Departamento de la UTI (Unidad de Telecomunicaciones e Investigación) de la Universidad Nacional de Loja. Se obtuvo respuesta bajo el tiquete de seguimiento GLPI #0017269. Los cuales hacen entrega de los datos en formato csv, constando con un total de 36363 registros académicos con 51 atributos de cada uno de los estudiantes pertenecientes a los años 2007 – 2020 (Ver Figura 1.).

1	facultad	carrera	periodo_lect	nivel	modalidad	cedula	apellidos	nombres	fecha_nacim	genero	estado_civil	etnia	sector_proce
2	AEIRNNR	COMPUTACI	2017-2018	DE GRADO	R presencial	706592268	ABALO PALA	JEAN PIERRE	8/14/2000	masculino	Soltero(a)	Mestizo	Urbano
3	AEIRNNR	COMPUTACI	2017-2018	DE GRADO	R presencial	706592268	ABALO PALA	JEAN PIERRE	8/14/2000	masculino	Soltero(a)	Mestizo	Urbano
4	AEIRNNR	COMPUTACI	2017-2018	DE GRADO	R presencial	706592268	ABALO PALA	JEAN PIERRE	8/14/2000	masculino	Soltero(a)	Mestizo	Urbano
5	AEIRNNR	COMPUTACI	2017-2018	DE GRADO	R presencial	706592268	ABALO PALA	JEAN PIERRE	8/14/2000	masculino	Soltero(a)	Mestizo	Urbano
6	AEIRNNR	COMPUTACI	2017-2018	DE GRADO	R presencial	706592268	ABALO PALA	JEAN PIERRE	8/14/2000	masculino	Soltero(a)	Mestizo	Urbano
7	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	151464468	ACHUPALLAS	MARVIN FAB	3/11/2002	masculino	Soltero(a)	Mestizo	Urbano
8	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	151464468	ACHUPALLAS	MARVIN FAB	3/11/2002	masculino	Soltero(a)	Mestizo	Urbano
9	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	151464468	ACHUPALLAS	MARVIN FAB	3/11/2002	masculino	Soltero(a)	Mestizo	Urbano
10	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	151464468	ACHUPALLAS	MARVIN FAB	3/11/2002	masculino	Soltero(a)	Mestizo	Urbano
11	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	151464468	ACHUPALLAS	MARVIN FAB	3/11/2002	masculino	Soltero(a)	Mestizo	Urbano
12	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
13	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
14	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
15	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
16	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
17	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
18	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
19	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
20	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano
21	AEIRNNR	COMPUTACI	2019-2020	DE GRADO	R presencial	1150472668	AGREDA SAR	JEAN XAVIER	4/1/2002	masculino	Soltero(a)	Mestizo	Urbano

Figura 45. Datos originales de los registros académicos

Dentro del conjunto se datos de puede visualizar muchos registros vacíos dentro de cada atributo; así mismo, el dataset cuenta con información de variables categóricas, numéricas, etc. Para obtener una buena data es necesaria la conversión de ciertas variables acorde a la técnica de minería de datos seleccionada, por ello se ha realizado un diccionario el cual permitirá identificar el significado de valores que tomen las variables dentro del data set.

## 5. Implementación

Para la aplicación de la Minería de Datos al caso de estudio se utilizó la metodología CRISP-DM, ya que permitió conseguir modelos descriptivos de minería de datos para identificar información relevante dentro de los registros académicos de los estudiantes. La metodología consta de las siguientes fases, estas se detallan a continuación.

### 5.1. FASE I. Comprensión del Negocio

Esta fase se centra en comprender cada uno de los objetivos y los requerimientos necesarios para la elaboración y ejecución del TT desde una perspectiva del negocio, con el propósito general de encontrar una solución que permita cumplir con cada uno de ellos de manera satisfactoria.

#### 5.1.1. Determinar los objetivos empresariales

La deserción de estudiantes en instituciones de educación superior ha representado un gran inconveniente no solo para la institución,



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

sino que también para las autoridades de gobierno, debido al alto número de abandono académico durante la formación del estudiante. Ante esto la Universidad Nacional de Loja no se aleja de la situación que se vive, es por ello que a través de una de ramas de la Inteligencia Artificial como lo es la Minería de datos Educativa, se desea encontrar los patrones existentes dentro de los registros académicos haciendo hincapié en asignaturas relacionadas con las matemáticas con estudiantes de Ingeniería, esto considerando que los altos números de abandono se da más en Carreras STEM (Ciencia, Tecnología, Ingeniería y Matemáticas).

### **5.1.2. Evaluar la situación Inicial**

Dentro de la UNL en cada una de sus Carreras ofertadas existe abandono estudiantil, sin embargo, esto se presenta más en Carreras de Ingeniería. Las autoridades correspondientes buscan soluciones para aumentar el número de estudiantes dentro de estas Carreras, así como también su permanencia y culminación. Los datos proporcionados por el departamento de Telecomunicaciones de la Universidad están presentes desde el año 2007 hasta la actualidad (Marzo 2020), pero cabe recalcar que debido a que el SGA en sus inicios mantenía un sistema modular, dentro de este no existía la automatización de ciertos procesos, los cuales fueron implementados desde en el nuevo sistema desde el periodo 2012-2013, es por eso que se muestra cierta inconsistencia en los datos en esas fechas.

### **5.1.3. Establecer los objetivos de minería de datos**

Para el presente TT se pretende identificar cuáles son los patrones que más predominan en la deserción de estudiantes a través de técnicas de minería de datos educativa, para ello se debe hacer la selección de la herramienta de minería de datos a utilizar con la que se va llevar a cabo el desarrollo de la investigación. De la misma manera se busca determinar el algoritmo más óptimo y realizar un análisis de los resultados obtenidos con el uso de las técnicas seleccionadas.

#### **5.1.4. Redactar el plan de proyecto**

Para poder cumplir con los objetivos establecidos se ha hecho uso de la Metodología CRIP-DM y para su cumplimiento se plantean lo siguiente: Recolectar los datos históricos a través del departamento encargado dentro de la Universidad Nacional de Loja, para ello se ha realizado la petición correspondiente a través de una solicitud enviada por medio de Gmail. Una vez obtenida respuesta se procede a trabajar con los datos históricos de los estudiantes de las Carreras de Ingeniería en Sistemas/Computación. Para la comprensión del dataset en general y de cada una de sus variables, se realiza un análisis exploratorio de los mismos, de este se obtendrá un informe técnico con cada una de las variables presentes dentro del dataset original; para el procesamiento y limpieza de los datos se lo llevará a cabo bajo la implementación de algoritmos en Python haciendo uso de las librerías, esto en conjunto con el uso de otras herramientas necesarias con base al tema de estudio. Una vez comprendido como se encuentra la data con la que se va a trabajar se procederá a aplicar los algoritmos acordes a la misma, y de esta manera encontrar la información predominante dentro de estos registros académicos, y con ello elaborar un plan de acción que sirva de apoyo a las autoridades correspondientes en la toma de decisiones y a su vez tratar de mitigar la deserción estudiantil dentro de la UNL.

### **5.2. FASE II. Comprensión de los Datos**

En esta fase se llevó a cabo la recolección de los registros académicos de los estudiantes de la Carrera de Ingeniería en Sistemas/Computación para una mejor comprensión de estos, ya que es la primera etapa para posteriormente realizar un análisis minucioso y con ello identificar como se encuentran estructurados los datos, este proceso se realizó a través de Excel en formato .csv, permitiendo obtener una mejor comprensión de cada una de las variables.

#### **5.2.1. Recoger datos Iniciales**

Los datos recolectados pertenecen a los estudiantes que han sido y que forman parte de la Carrera de Ingeniería en



Sistemas/Computación de la Universidad Nacional de Loja, concernientes a los años 2007-2020. Los datos en estado bruto se encuentran en un archivo .csv, este cuenta con 36363 registros y 51 variables como se muestran en la Figura 1. Consta de variables sociales, demográficas, económicas y académicas de cada uno de los estudiantes a manera general.

### 5.2.2. Describir Datos

La Tabla XIII muestra cada una de las variables que se determinaron en el conjunto de datos, estos se detallan a continuación.

Tabla XIII. Variables dentro del conjunto de datos.

Variable	Descripción
facultad	Describe la Facultad a la que pertenecen las Carreras de los estudiantes dentro de la UNL.
carrera	Describe el nombre de la Carrera a la que corresponde el estudiante.
periodo_lectivo	Nombre que se le asigna a la fecha del ciclo académico.
nivel	Representa el régimen de los estudiantes dentro de la universidad.
modalidad	Modalidad de estudios de las Carreras
cedula	Cédula de identidad de los estudiantes
apellidos	Representa los apellidos de los estudiantes
nombres	Representa los nombres de los estudiantes
fecha_nacimiento	Contiene la fecha de nacimiento de los estudiantes
genero	Genero de los estudiantes (Masculino o Femenino)
estado_civil	Estado civil de los estudiantes
Etnia	Esta variable representa la etnia de los estudiantes
sector_procedencia	Describe el sector de procedencia de los estudiantes
nacionalidad	Describe el estado al que pertenece el estudiante
pais_nacimiento	Describe el país de nacimiento de los estudiantes



provincia_nacimiento	Describe la provincia de nacimiento de los estudiantes
canton_nacimiento	Describe el cantón de nacimientos de los estudiantes
ciudad_nacimiento	Describe la ciudad de nacimiento de los estudiantes
direccion_nacimiento	Describe la dirección de nacimiento de los estudiantes
pais_actual	Describe el país actual de los estudiantes
provincia_actual	Describe la provincia actual de los estudiantes
canton_actual	Describe el cantón actual de los estudiantes
ciudad_actual	Describe la ciudad actual de los estudiantes
direccion_actual	Describe la dirección actual de los estudiantes
trabaja	Describe si los estudiantes trabajan o no.
ingreso_estudiante	Describe el monto total de ingresos de los estudiantes
numero_hijos	Describe el número de hijos por parte de los estudiantes
colegio	Describe a la institución de formación de los estudiantes
tipo_colegio	Describe el tipo de institución de los estudiantes, es decir, público, privado o fisco misional.
pais_colegio	Describe el país donde se encuentra el colegio de los estudiantes
provincia_colegio	Describe la provincia de colegio de los estudiantes
canton_colegio	Describe el cantón de colegio de los estudiantes
oferta_academica	Describe la oferta académica para el ciclo académico
ciclo	Describe el ciclo que cursan los estudiantes
numero_matricula	Describe el número de matrícula de los estudiantes
malla_curricular	Describe la malla a la que pertenecen los estudiantes
estado_matricula	Describe el estado de matrícula de los estudiantes





paralelo	Describe el paralelo al que pertenecen los estudiantes
jornada	Muestra la jornada en la que los estudiantes reciben clases
asignatura	Describe a cada una de las asignaturas que deben aprobar los estudiantes
estado_asignatura	Representa el estado de la asignatura de los estudiantes es decir si el alumno aprueba o reprueba.
obligatoria	Describe el estado obligatoriedad de las asignaturas.
arrastrable	Describe si la asignatura que cursan los estudiantes es arrastrable o no.
asistencia_obligatoria	Describe la asistencia obligatoria de las asignaturas por parte de los estudiantes.
nota_ingresada	Describe la calificación de la asignatura ingresada cursada por los estudiantes.
nota_ponderada	Describe el 20% de la nota ingresada por cada asignatura.
porcentaje_asistencia	Describe el % de asistencia a cada asignatura por los estudiantes
promedio_matricula	Describe el promedio de la matrícula de los estudiantes
homologacion	Describe si la asignatura cursada por el estudiante es homologada o no.
observacion_homologacion	Describe si existe alguna observación dentro del proceso de homologación

### 5.2.3. Explorar Datos

Para poder conseguir una estructura general de los datos, se realizó un análisis estadístico con Python, lo cual permitió obtener un conocimiento más a fondo de la data inicial, sus variables, los valores inconsistentes, valores perdidos (missing), fuera de rango (atípicos) y en algunos casos no definidos. Esto se ve reflejado en el **Anexo 2.**

### **Análisis y Exploración de los datos**

### 5.2.4. Verificar la calidad de los datos

Dentro del conjunto de datos para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores vacíos o nulos, y para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso se ha verificado

cada una de las variables, revisando su estado y las características que presentan, esto nos permite identificar la confiabilidad de las variables obtenidas.

### 5.3. FASE II. Preparación de los Datos

Para la ejecución de esta fase se llevó a cabo la preparación de los datos con los cuales se puedan trabajar de acuerdo a las técnicas de minería de datos seleccionadas, para ello fue necesaria la estructura de datos de la fase anterior, del mismo modo, se llevó a cabo una limpieza de los datos para poder generar el modelo de minería de datos, tratando de obtener una data sin valores nulos o vacíos.

Así mismo, con la estructura de datos seleccionada en base a variables **cedula, fecha\_nacimiento** se generaron nuevos atributos a partir de los que ya existían, por otra parte, se realizó agrupación de campos con características comunes para finalmente transformar los datos sin que exista modificaciones en su significado, de esta manera se ajustan a las técnicas de minería de datos arrojando mejores resultados.

#### 5.3.1. Selección de Datos

Para trabajar dentro del contexto de la investigación y basándose en [1], [3]–[7], [27], [121], fue necesario eliminar algunas columnas que no aportaban ningún valor para cumplir con los objetivos de esta, entre ellas tenemos:

- facultad
- periodo\_lectivo
- nivel
- modalidad
- apellidos
- nombres
- nacionalidad
- país\_nacimiento
- canton\_nacimiento
- ciudad\_nacimiento
- dirección\_nacimiento
- país\_actual
- provincia\_actual
- canton\_actual
- ciudad\_actual
- direccion\_actual
- colegio
- país\_colegio
- provincia\_colegio
- canton\_colegio
- oferta\_académica
- numero\_matricula
- paralelo
- jornada
- obligatoria
- arrastrable
- asistencia\_obligatoria
- nota\_ingresada
- nota\_ponderada
- porcentaje\_asistencia
- promedio\_matricula
- homologación
- observación\_homologación

En base a cada una de las actividades dentro de las fases anteriores y en base al objetivo de la investigación las variables que se usarán para la minería de datos en base al investigador son las que se detallan en la Tabla XIV:

Tabla XIV. Variables para la minería de datos

Variable	Descripción	Observación
carrera	Nombre de la Carrera	Esta variable si se utilizara considerando: Sistemas = 1 Computación = 0
cedula	Cédula de identidad de los estudiantes	Esta variable no será utilizada en el dataset, pero se generará un id, con el secuencial de las cédulas únicas.
fecha_nacimiento	Fecha de nacimiento de los estudiantes	Esta variable directamente no servirá para el dataset, pero será utilizada para obtener la edad de los estudiantes
genero	Genero de los estudiantes	Esta variable si será utilizada en el dataset considerando: Hombres = 0 Mujeres = 1
estado_civil	Estado civil de los estudiantes	Esta variable si será utilizada en el dataset.
Etnia	Esta variable representa la etnia de los estudiantes	Esta variable si será utilizada en el dataset.
sector_procedencia	Describe el sector de procedencia de los estudiantes	Esta variable si será utilizada en el dataset.
trabaja	Describe si los estudiantes trabajan	Esta variable si será utilizada en el dataset
ciclo	Describe el ciclo que cursan los estudiantes	Esta variable si será utilizada en el dataset para identificar los patrones de acuerdo a cada ciclo.
asignatura	Describe a cada una de las asignaturas que	Esta variable si será utilizada en el dataset considerando que también se hará uso de

	deben aprobar los estudiantes	las asignaturas relacionadas a las matemáticas de acuerdo con el código de la UNESCO.
estado_asignatura	Describe el estado de la asignatura de los estudiantes	Esta variable si será utilizada en el dataset de acuerdo con cada asignatura, buscando relación entre ellas considerando: Reprobado = 0 Aprobado = 1
provincia_nacimiento	Describe la provincia de la cual es oriundo el estudiante	Esta variable se la utilizará para identificar si las personas que provienen de otras provincias destacan más que lo que pertenecen a la provincia de Loja
tipo_colegio	Describe la Institución de la que proviene el estudiante	Esta variable servirá para tratar si la procedencia de una Institución Fiscal, Fiscomisional o Privada, influyen en el desenvolvimiento del estudiante durante los primeros ciclos de tu formación académica. Fiscal = 1 Fiscomisional = 2 Privada = 3

Adicionalmente, a nivel general, la Figura 46 muestra la estructura con la que se agrupó a cada una de las variables del conjunto de datos, desde la recepción de los datos, la categorización de los mismo y finalmente, el producto final con el que se va a trabajar.

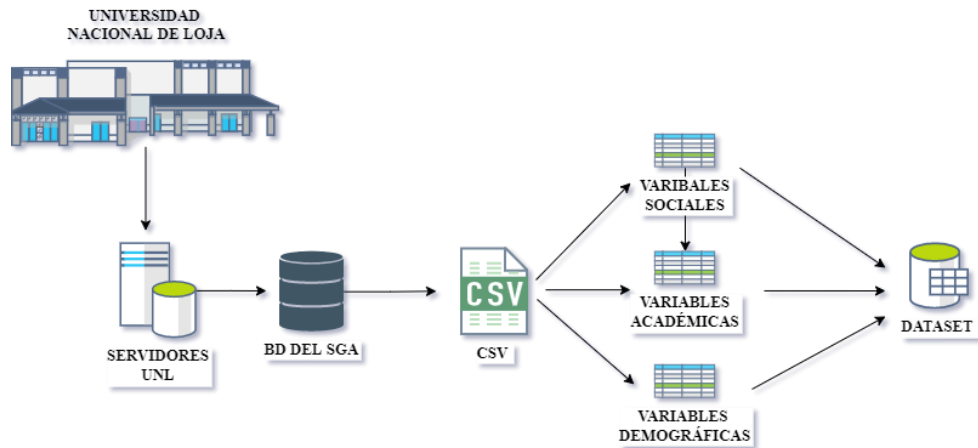


Figura 46. Estructura del conjunto de datos

### 5.3.2. Limpiar Datos

La limpieza de los datos se la realizó en primera instancia, eliminando las variables que de acuerdo con el objeto de estudio no son representativas, para otras que resultaban desconocidas se procedió a solicitar el apoyo de un experto el cuál orientó en la identificación de ciertos parámetros necesarios para la continuidad del proceso de limpieza; para otros atributos como el tipo de colegio en donde se necesitaban predecir registros a partir de los existentes, se lo realizó mediante el *algoritmo K-NN*, expuesto en ***relleno\_datos\_1.py***. Una vez obtenido el dataset y a través de Spyder un IDE de Python, se procedió a eliminar los registros que no contaban con la información necesaria en relación a todas las variables que anteceden, posteriormente se procedió a clasificar por Carreras a través del algoritmo denominado ***limpieza\_general\_2.py***, con el cual se obtuvo un total de 29 005 registro para Sistemas y 2 449 registros para computación, seguidamente, a través de Excel y de forma manual se procedió a separar los dataset en ciclos. Para aquellos registros académicos faltantes relacionados con las asignaturas se basó en [122]–[124] para aplicar la técnica de *imputación por media* y de esta manera completar a los mismos plasmándose a través del algoritmo ***limpieza\_especifica\_3.py***.

Todos los algoritmos, así como los conjuntos de datos obtenidos se encuentran almacenados en el repositorio de **GitHub**<sup>8</sup>.

### 5.3.3. Construir datos

Una vez finalizada la limpieza de los datos, se logró obtener un conjunto definitivo para poder aplicar las técnicas de minería de datos específicas.

Para ello se ha elaborado un conjunto de estados (Ver Tabla XVI y XVII) el mismo que permitirá reconocer el valor de las variables que se obtengan de los resultados.

Tabla XV. Diccionario de variables sistemas/computación

Genero		Provincia de nacimiento	
Mujer	1	Aprobada	1
Hombre	0	Reprobada	0
Estado civil		Tipo de colegio	
Soltero(a)	0	Fiscal	1
Casado(a)	1	Fiscomisional	2
Sector de Procedencia		Privado	3
Rural	1	Etnia	
Urbano	0	Mestizo	1
Trabaja		Blanco	2
Si	1	Indígena	3
No	0	Mulato	4
Estado Asignatura		Negro	5
Aprobada	1	Montubio	6
Reprobada	0		

Tabla XVI. Codificación para las provincias de nacimiento sistemas/computación

Provincia De Nacimiento	
Exterior	0
Azuay	1
Bolívar	2
Cañar	3

<sup>8</sup> [https://github.com/Jhulissa/data\\_mining\\_for\\_student\\_dropouts](https://github.com/Jhulissa/data_mining_for_student_dropouts)



Carchi	4
Cotopaxi	5
Chimborazo	6
El Oro	7
Esmeraldas	8
Guayas	9
Imbabura	10
Loja	11
Los Ríos	12
Manabí	13
Morona Santiago	14
Napo	15
Pastaza	16
Pichincha	17
Tungurahua	18
Zamora Chinchipe	19
Galápagos	20
Sucumbíos	21
Orellana	22
Santo Domingo De Los Tsáchilas	23
Santa Elena	24

Para la codificación de las asignaturas, se ha considerado la primera letra de cada palabra, por ejemplo: Cálculo Diferencia = CD, Inteligencia Artificial = IA, etc. Esto para cada una de las materias que se presentan por ciclo.

Asignados estos valores necesarios para la construcción del dataset final, se utilizó el algoritmo **limpieza\_especifica\_3.py** el cual permitió la transposición de las asignaturas que estaban dentro de una columna en filas, en relación a cada estudiante tal y como se indica en la figura 2 (para ello fue necesario mantener el número de cédula, una vez culminado se procede a eliminarlo) con ello se logra

tener un dataset con todos los requerimientos para aplicar la minería de datos.



cedula	edad	genero	estado	etnia	sector_trabaja	ciclo	asignatura	estado_asignatur	id	CALCULO_DI	FUNDAMENT	PROGRAMAC	CALCULO_IN	PROBABILID	ALGEBRA_LI	EST
1150137642	25	1	1	1	1	0	1 ALGEBRA LINEAL	0	2	0	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	2 QUIMICA	1	3	1	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	3 EXPRESION ORAL	0	4	2	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	4 EXPRESION ORAL	1	5	3	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	5 CALCULO DIFEREN	1	6	4	2	2	2	2	2	2
1150137642	25	1	1	1	1	0	6 TEORIA DE LA PRC	0	7	5	2	2	2	2	2	2
1150137642	25	1	1	1	1	0	7 FISICA	1	8	6	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	8 FUNDAMENTOS II	1	9	7	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	9 LABORATORIO DE	0	10	8	1	1	0	1	1	1
1150137642	25	1	1	1	1	0	10 CALCULO DIFEREN	0	11	9	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	11 FISICA	0	12	10	1	1	0	0	0	0
1150137642	25	1	1	1	1	0	12 FUNDAMENTOS II	0	13	11	0	0	2	2	2	2
1150137642	25	1	1	1	1	0	13 PROGRAMACION	1	14	12	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	14 CALCULO INTEGR	1	15	13	0	0	1	1	1	1
1150137642	25	1	1	1	1	0	15 PROBABILIDAD E	1	16	14	0	1	1	1	0	1
1150137642	25	1	1	1	1	0	16 ECOLOGIA Y MED	1	17	15	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	17 ALGEBRA LINEAL	1	18	16	0	1	0	2	0	0
1150137642	25	1	1	1	1	0	18 ESTRUCTURA DE I	1	19	17	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	19 FISICA II	1	20	18	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	20 MATEMATICAS DI	1	21	19	1	1	1	1	1	1
1150137642	25	1	1	1	1	0	21 ESTRUCTURA DE I	0	22	20	2	2	2	2	2	2

Figura 47. Transposición de asignaturas únicas a filas

### 5.3.4. Integrar Datos

Dentro del conjunto de datos esta actividad no fue desarrollada puesto que no fue necesario elaborar una base de datos o relacionar distintas tablas con diversos contenidos ya que, la data proporcionada por la UTI fue enviada de manera general y con toda la información requerida para el desarrollo del TT.

### 5.3.5. Formatear Datos

Esta tarea no fue desarrollada por el investigador ya que en base a pruebas se logró identificar mejores resultados dejando tal y como se encuentra.

## 5.4. FASE IV. Modelado

Dentro de esta fase se llevó a cabo la selección de la técnica de modelado que más se asemeje al TT.

### 5.4.1. Seleccionar técnicas de modelado

Para la selección de la técnica de Minería de Datos acorde al TT se ha basado en [21], [125], [126] [43], y al mismo tiempo en los resultados obtenidos a través de la RSL, con ello se logró seleccionar las técnicas de Clustering y Reglas de asociación, con algoritmos de Simple K-means y A priori respectivamente, para búsqueda de patrones dentro de la información proporcionada y obtenida.

### 5.4.2. Diseñar las pruebas del modelo

Para el plan de pruebas se empleará el uso de métricas que permitirán medir o indicar a las reglas de asociación más relevantes



e importantes para el tema en estudio, para ello, de acuerdo con [127] se describe:

La **cobertura (coverage)**, denominada completitud horizontal, establece cuántos de todos los posibles casos observables del mundo real están representados por los datos disponibles. La granularidad es necesaria a nivel de conjunto de datos y su valor suele estar representados en porcentaje.

La **confianza** o confiabilidad, indica que tan confiables son los datos, es decir, que tan creíbles resultan para el usuario, esta se representa con valor de 0 o 1 y en ciertas herramientas con valores porcentuales, donde 1 indica que la regla es 100% confiable.

**Elevación (Lift)**, calcula la relación entre X e Y (realmente se entiende como la confianza de la regla dividida por el soporte del consecuente de la regla). Esta métrica obtiene valores iguales, mayores o menores que 1, con los que se identifica la forma de la relación: si  $lift > 1$  = X e Y correlacionan positivamente; si  $lift < 1$  = X e Y correlacionan negativamente; si  $lift = 1$  = X e Y son independientes. Entre mayor sea el valor, más alta es la probabilidad de que la existencia de una transacción  $X \rightarrow Y$  no sea aleatoria

**Distancia en puntos**, métrica utilizada por el algoritmo K-means para ver la distancia entre los datos.

Cabe señalar que la evaluación de los modelos se lo va a llevar a cabo por separado, es decir, por cada uno de los ciclos, tanto en Sistemas como en Computación.

#### **5.4.3. Construir los modelos**

Para el cumplimiento de esta actividad se ha considerado únicamente ciertas variables demográficas, sociales y académicas en donde, se ha seleccionado únicamente las asignaturas

relacionadas a las matemáticas, esto de acuerdo al código de la UNESCO, además, se ha omitido ciertas variables que han demostrado irrelevancia para el modelo. Considerar que la aplicación de los algoritmos es igual para cada uno de los modelos obtenidos por ciclo, cada uno de ellos con sus respectivas características y atributos.

Los algoritmos seleccionados se los ha ejecutado dentro de la herramienta Weka, y como sustento a esta, se ha implementado también dentro de la herramienta RapidMiner y en Python, mismas que fueron seleccionadas de acuerdo a la RSL y se detallan a continuación:

Para la elaboración del modelo a nivel general en la herramienta **Weka**, primeramente, se carga el archivo .arff propio de este o a su vez un .csv con la data de todos los atributos, a continuación, se muestra el proceso:

Al ejecutar Weka se muestra una ventana como se ve en la figura 3.

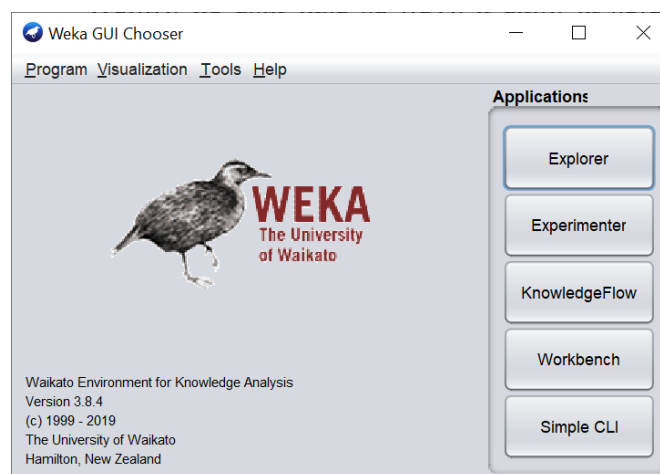


Figura 48. Interfaz principal de Weka

Dentro de la interfaz seleccionamos explorar, al desplegarse una nueva ventana, en la parte superior izquierda se muestra un ícono *Open file* esto permitirá subir el archivo para ejecutarse (Ver Figura 4).

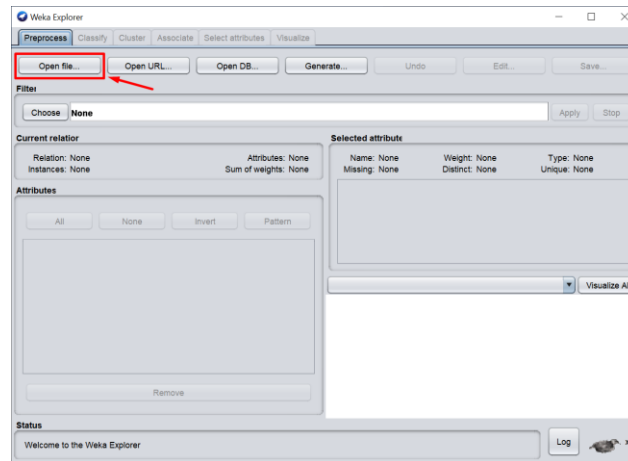


Figura 49. Interfaz de Weka para subir archivos a procesar.

En **RapidMiner**, en cambio, se muestra una pantalla que permite modelar desde cero todo el algoritmo, tal y como lo indica la Figura 8. Dentro de ella se procede a generar o diseñar el modelo cargando el dataset y todas las funciones que sean necesarias expuestas dentro de *operators*.

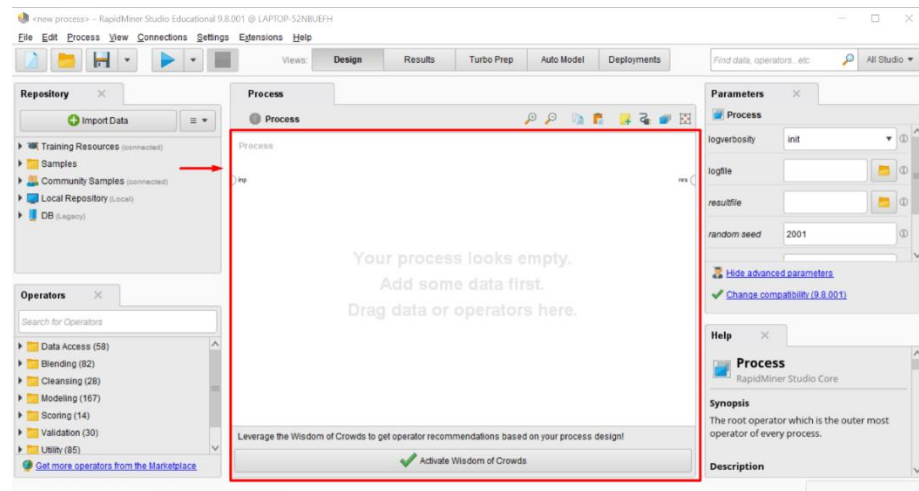


Figura 50. Interfaz de RapidMiner

A continuación, a nivel específico muestra paso a paso la construcción del modelo a nivel de cada herramienta.

#### 5.4.3.1. Algoritmo A priori - Weka

Como se indicó con anterioridad, se procedió a elaborar el modelo de acuerdo con cada herramienta. En Weka, para las reglas de asociación, haciendo uso del Algoritmo A priori y cargado el dataset,

se procede a ingresar a *Associate* donde se encuentra el algoritmo A priori como lo indica la Figura 51.

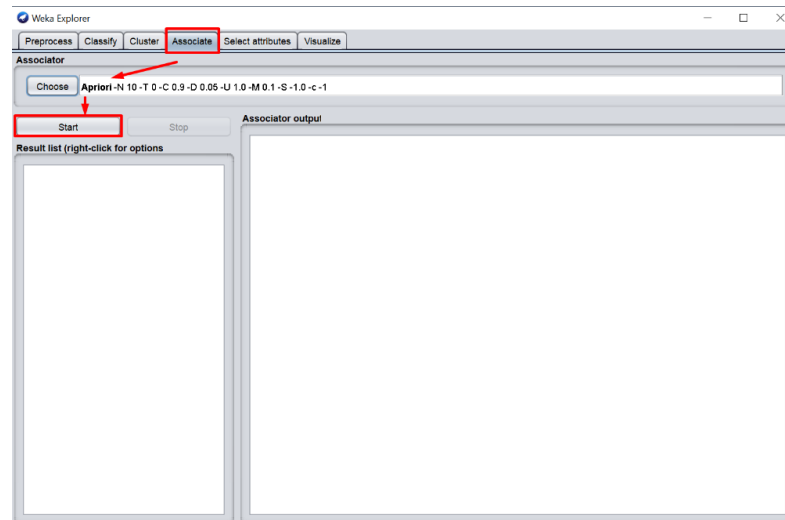


Figura 51. Crear reglas de asociación

Para configurar el número de reglas y la métrica a utilizar se procede a dar clic en A priori y se desplegará una ventana en donde se realiza dicha configuración. A continuación, se da clic en start, con ello se inicializa el proceso y se obtienen las reglas de asociación.

#### 5.4.3.1. Algoritmo A priori en RapidMiner

Para este algoritmo, como ya se tiene cargado previamente el dataset (Ver Figura 52), se procede a añadirle los componentes necesarios para dicho algoritmo descritos a continuación:

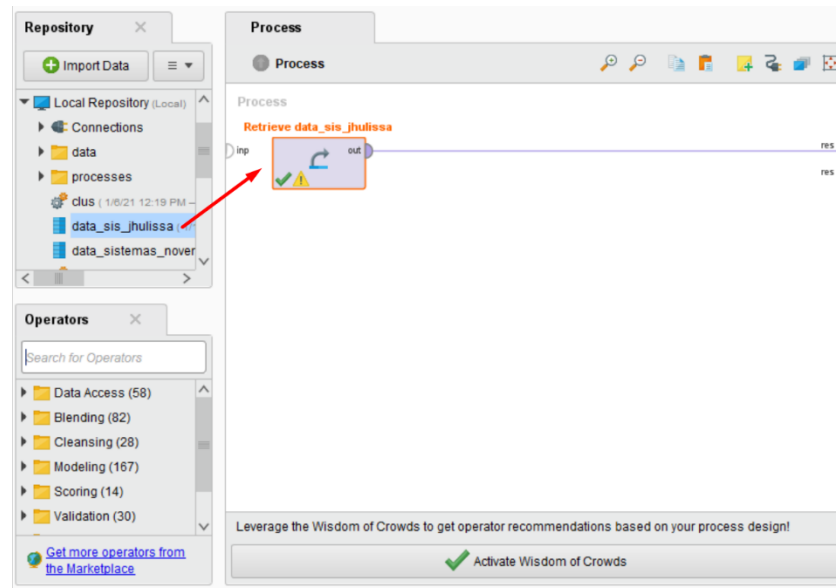
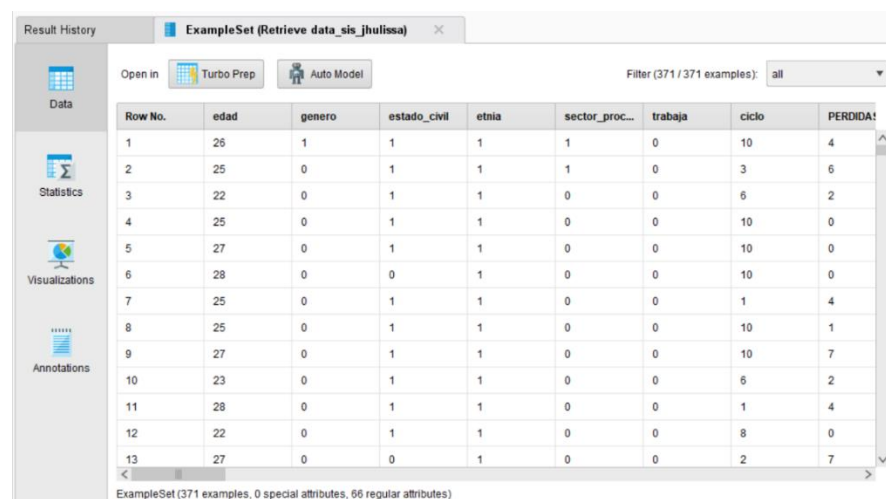


Figura 52. Cargar dataset en RapidMiner

Al ejecutar el dataset cargado dentro de la herramienta (Ver Figura 53) se desplaza todas las instancias con los atributos presentes en el conjunto de datos, cada uno con sus respectivos valores.



Row No.	edad	genero	estado_civil	etnia	sector_proc...	trabaja	ciclo	PERDIDA
1	26	1	1	1	1	0	10	4
2	25	0	1	1	1	0	3	6
3	22	0	1	1	0	0	6	2
4	25	0	1	1	0	0	10	0
5	27	0	1	1	0	0	10	0
6	28	0	0	1	0	0	10	0
7	25	0	1	1	0	0	1	4
8	25	0	1	1	0	0	10	1
9	27	0	1	1	0	0	10	7
10	23	0	1	1	0	0	6	2
11	28	0	1	1	0	0	1	4
12	22	0	1	1	0	0	8	0
13	27	0	0	1	0	0	2	7

Figura 53. Conjunto de datos cargados

Para obtener las reglas de asociación se debe considerar el formato de los datos, es este caso se debe utilizar el componente Numerical to Polynomial, como se muestra en la Figura 54, esto nos permitirá convertir los valores tales como la cédula y las mismas asignaturas a polinomiales, debido a que no están construidas en base a valor

positivo y negativo si no que, es necesario una tercera opción para el procesamiento de los datos.

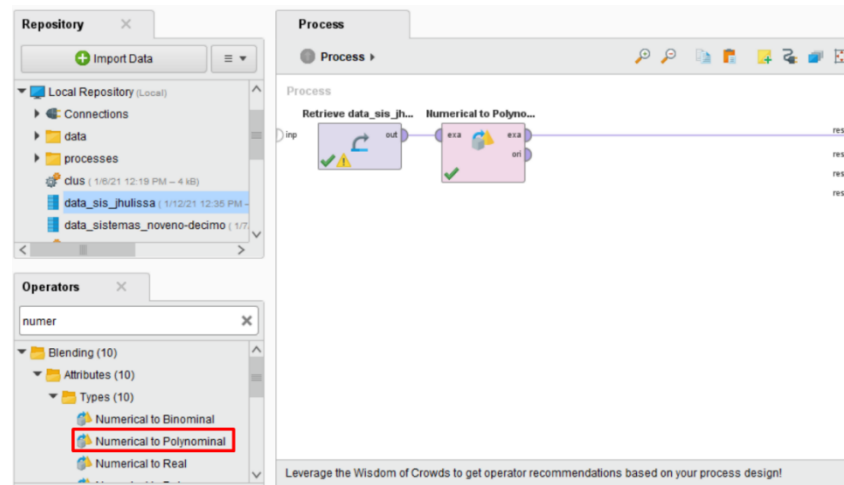


Figura 54. Implementar componente para transformar datos

El algoritmo A priori permite encontrar de forma eficiente conjuntos de ítems frecuentes, los cuales sirven de base para generar reglas de asociación. Para ello es necesario cargar el componente FP-Growth (Ver Figura 55) encargado de calcular los elementos que ocurren con frecuencia en una base de datos de transacciones mostrando una tabla de combinaciones posibles con un soporte mínimo del 0.6 (Ver Figura 56)

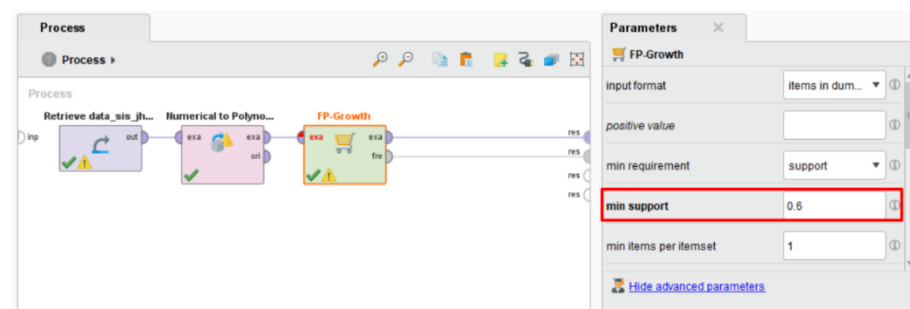


Figura 55. Añadir componente FP-Growth

Size	Support ↓	Item 1	Item 2	Item 3
1	0.916	sector_procedencia		
1	0.790	genero		
2	0.725	sector_procedencia	genero	
1	0.722	PROYECTOS INFORMATICOS II		
2	0.658	sector_procedencia	PROYECTOS INFORMATICOS II	
2	0.571	genero	PROYECTOS INFORMATICOS II	
3	0.520	sector_procedencia	genero	PROYECTOS INFORMATICOS II

Figura 56. Resultados obtenidos por componente FP-Growth

Finalmente, agregamos un componente que genere las reglas de asociación denominado Create Association Rules, con un número de confianza principal del 0.6, como se muestra la Figura 57

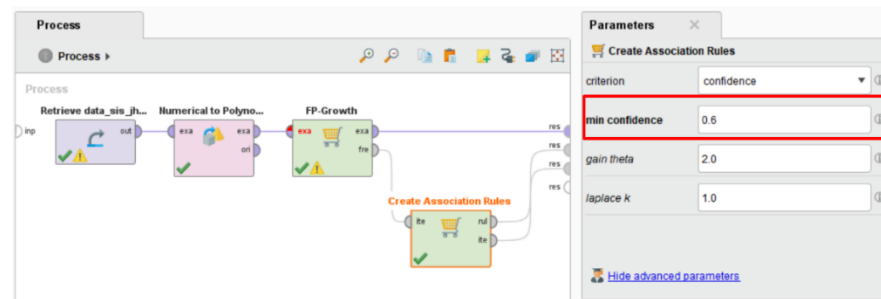


Figura 57. Agregar Componente para Reglas de Asociación.

## Algoritmo K-means - Weka

Cargado el archivo arff a Weka y dentro de la sección Choose seleccionamos el algoritmo en este caso dentro de Clúster a Simple K-means, tal y como se muestra en la Figura 58.

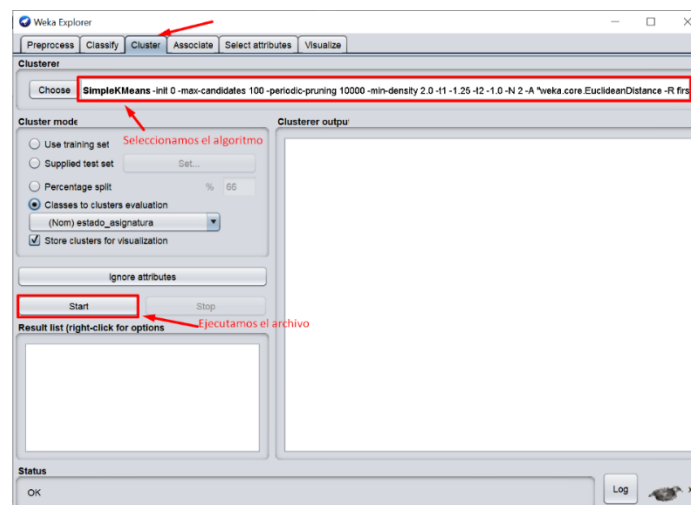


Figura 58. Ejecutar algoritmo Simple KMeans

Dentro de las propiedades del algoritmo se seleccionó el número de clústers que se desean obtener, en este caso 5 clústers (Ver Figura 59).

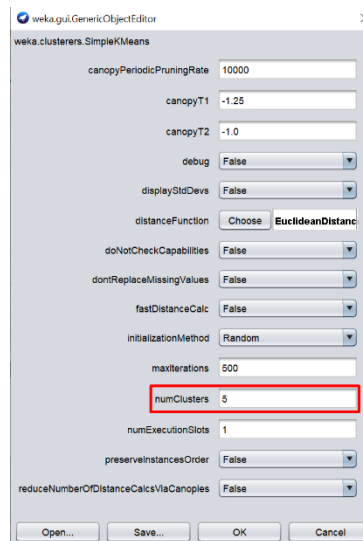


Figura 59. Configuración del Algoritmo KMeans

## Algoritmo K-means – RapidMiner

Para este algoritmo dentro de esta herramienta se procede a cargar el dataset para efectuar el algoritmo, tal y como se indica en la Figura 60.

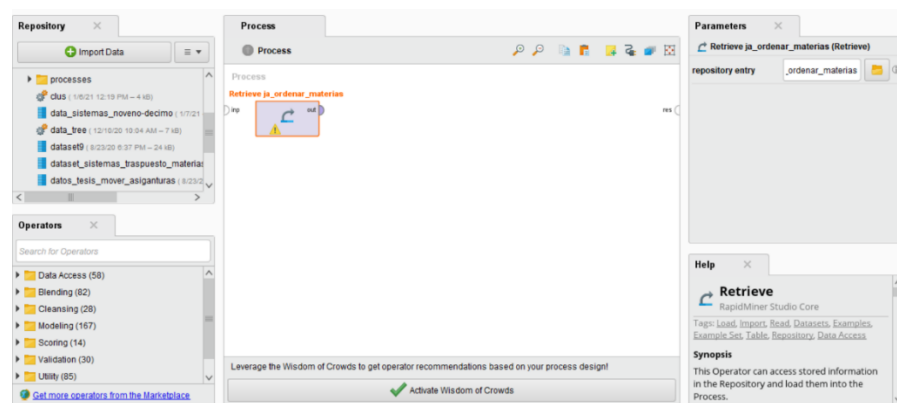


Figura 60. Algoritmo k-means en RapidMiner

Para ejecutar el algoritmo K-means es necesario un componente de Normalización (Ver Figura 61), el cual se utiliza para escalar los valores para que quepan en un rango específico. La normalización



es útil para comparar atributos que varían en tamaño. Este operador realiza la normalización de los atributos seleccionados.

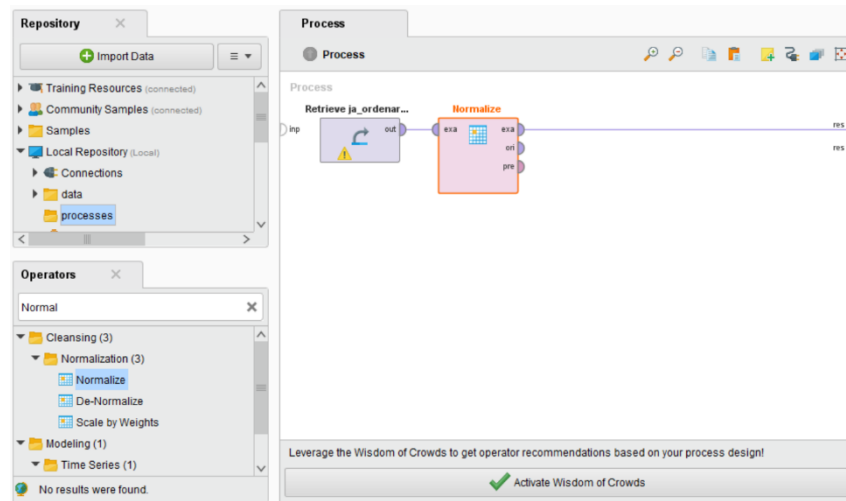


Figura 61. Añadir Componente de Normalización

Dentro de la normalización existen muchos métodos para transformar rangos, dentro de ellos se encuentra z-transformation, el cual consiste en restar la media y dividir para la desviación estándar, prestándose como la más apropiada para la normalización de los datos (Ver .

Añadido el componente de Normalización se procede a conectarlo con el algoritmo K-Means, presente en el algoritmo de clúster, tal y como se indica en la Figura...De la misma manera se procede a colocar el número de clústering a utilizar dentro del TT, en este caso se pondrán 5 clúster (Ver Figura 62)

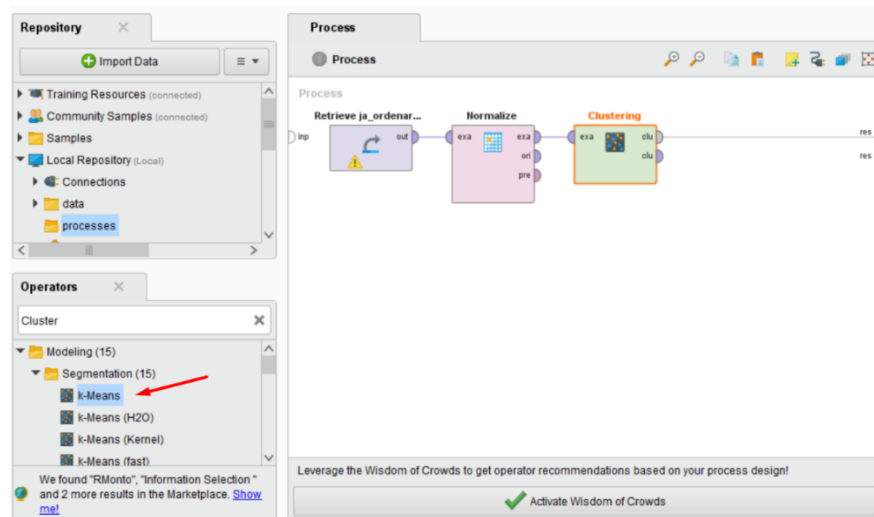


Figura 62. Añadir Componentes Normalización y KMeans

Finalmente se agrega el componente Performance, así como lo muestra la Figura 63, encargado de evaluar el desempeño del modelo, en relación a la distancia de cada clúster y a su centroide más cercano.

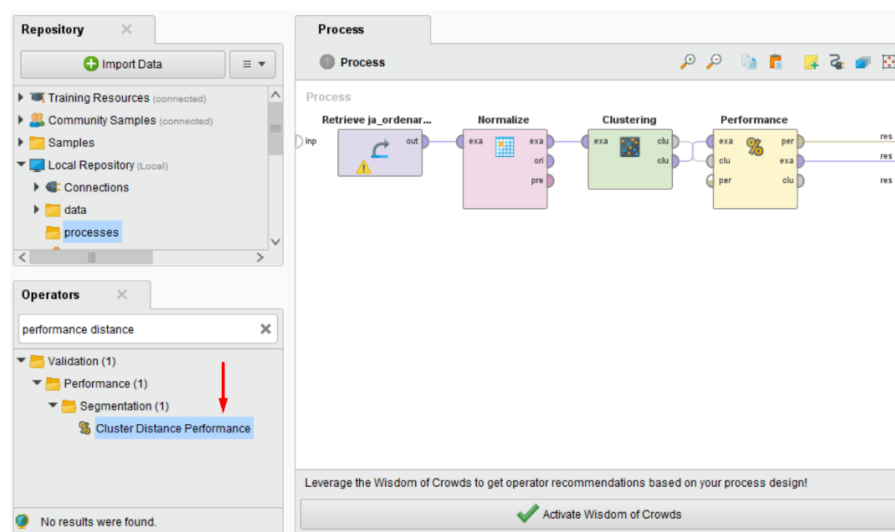


Figura 63. Agregar Componente Performance

Ejecutamos o corremos el modelo diseñado y en cada componente se cargará un visto como muestra que todo se ha realizado de manera correcta, así se muestra en la Figura 64

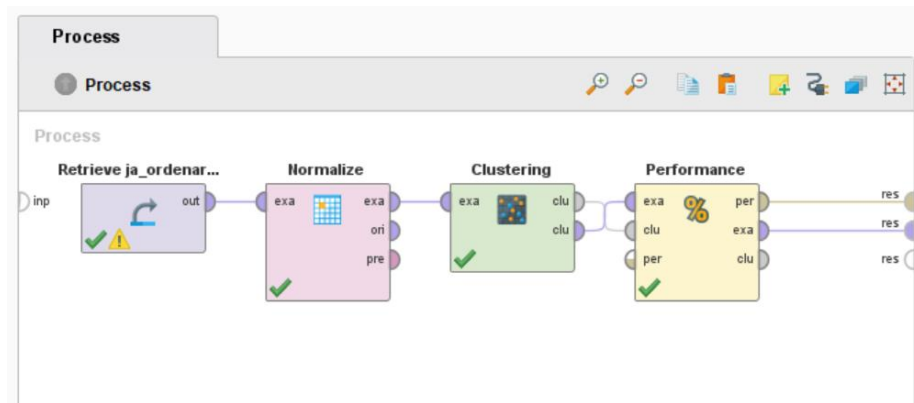


Figura 64. Diseño completo para Clústering

Adicionalmente, como se indicó en instancias anteriores hay atributos que no son binomiales y para ellos se debe añadir una de las funciones que permite mantener los datos como se deberían (Ver Figura 65), esta es Parse Numbers.



Figura 65. Componente adicional Parse Numbers

## Algoritmos A priori y K-means - Python

Para la elaboración y obtención de resultados de ambos algoritmos, se ha hecho uso de dos IDE's, Spyder y Google Colaboraty, la codificación de los mismos se encuentra almacenada en GitHub<sup>9</sup>

### 5.4.4 Evaluar los modelos

Una buena manera de evaluar la efectividad de los modelos es utilizando los indicadores o métricas que se establecieron en el plan de pruebas de este documento (Sección 5.4.2.), tanto para las reglas de asociación como para los clúster generados, para ello dentro de cada modelo generado y según la herramienta utilizada se obtienen los diferentes resultados correspondientes a cada ciclo de ambas Carreras, así como también, un informe

<sup>9</sup> [https://github.com/Jhulissa/data\\_mining\\_for\\_student\\_dropouts](https://github.com/Jhulissa/data_mining_for_student_dropouts)

sobre la obtención de los modelos en las tres herramientas utilizadas.

En virtud de lo anteriormente expuesto, el resultado de cada modelo obtenido se puede visualizar en cada uno de los Anexos correspondiente a cada ciclo.

Se debe considerar que se han realizado varias pruebas para llegar a los modelos finales, tratando de mantener la veracidad de los datos se ha trabajado bajo la técnica de prueba y error, esto permitió al investigador obtener los resultados con los que se trabajará el plan de acción.

A continuación, muestro una de los resultados que se obtuvieron mientras se aplicaba la técnica antes mencionada. En la Figura 66. Resultados de Prueba y Error se puede identificar que la relación entre los atributos no tiene ninguna relación ni representación, a pesar que los valores de la confianza parecían ser favores, no eran suficientes ni representativos para el objeto de estudio. En virtud de lo que antecede estos y muchos otros resultados fueron descartados por el Investigador al no aportar ningún valor al Proyecto de TT.

```
Best rules found:
1. LABORATORIO_DE_FISICA_I=2 315 ==> TEORIA_DE_LA_PROGRAMACION=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. TEORIA_DE_LA_PROGRAMACION=2 315 ==> LABORATORIO_DE_FISICA_I=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. LABORATORIO_DE_FISICA_II=2 315 ==> TEORIA_DE_LA_PROGRAMACION=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. TEORIA_DE_LA_PROGRAMACION=2 315 ==> LABORATORIO_DE_FISICA_II=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. LABORATORIO_DE_FISICA_II=2 315 ==> LABORATORIO_DE_FISICA_I=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. LABORATORIO_DE_FISICA_I=2 315 ==> LABORATORIO_DE_FISICA_II=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. LABORATORIO_DE_FISICA_I=2 LABORATORIO_DE_FISICA_II=2 315 ==> TEORIA_DE_LA_PROGRAMACION=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. TEORIA_DE_LA_PROGRAMACION=2 LABORATORIO_DE_FISICA_II=2 315 ==> LABORATORIO_DE_FISICA_I=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. TEORIA_DE_LA_PROGRAMACION=2 LABORATORIO_DE_FISICA_I=2 315 ==> LABORATORIO_DE_FISICA_II=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. LABORATORIO_DE_FISICA_II=2 315 ==> TEORIA_DE_LA_PROGRAMACION=2 LABORATORIO_DE_FISICA_I=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
11. LABORATORIO_DE_FISICA_I=2 315 ==> TEORIA_DE_LA_PROGRAMACION=2 LABORATORIO_DE_FISICA_II=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
12. TEORIA_DE_LA_PROGRAMACION=2 315 ==> LABORATORIO_DE_FISICA_I=2 LABORATORIO_DE_FISICA_II=2 315 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
13. etnia=1 305 ==> TEORIA_DE_LA_PROGRAMACION=2 305 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
14. etnia=1 305 ==> LABORATORIO_DE_FISICA_I=2 305 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
15. etnia=1 305 ==> LABORATORIO_DE_FISICA_II=2 305 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

Figura 66. Resultados de Prueba y Error

## 5.5. FASE V. Evaluación

En instancias anteriores, se estableció como criterio de éxito principal determinar los patrones dentro de los registros académicos de los estudiantes, que permitan conocer cuáles son los factores que influyen al abandono de su carrera, para lo cual es necesario tener una base objetiva,

como lo son las métricas o indicadores estadísticos obtenidos a través de la ejecución de los modelos.

Los resultados expuestos a continuación, se basan en cada uno de los modelos obtenidos por cada ciclo de ambas Carreras, estos se encuentran a detalle en el **Anexo 4. Ingeniería en Sistemas – Primer ciclo** hasta el **Anexo 17. Ingeniería en Computación – Cuarto ciclo**.

Dicho lo anterior y para una mejor comprensión se ha elaborado una tabla donde se identifica cual o cuales son las características más representativas en las reglas obtenidas de acuerdo a cada modelo generado y a las métricas o indicadores seleccionados, cada uno de ellas cuentan con su respectivo valor, esto de acuerdo al diccionario de datos elaborado (Ver Sección 5.3.3).

- **Primer ciclo – Sistemas**

Como primera instancia, haciendo referencia a variables, sociales y demográficas, tomando en consideración al primer ciclo y como indicador a la confianza, se identifica en la Figura 67 a estudiantes de género masculino, que no deben contar con ninguna actividad laboral, residir en el sector Urbano y la aprobación de la materia de Cálculo Diferencial, como las características más predominantes dentro del antecedente, mientras que, por el consecuente a aprobar la asignatura de Fundamentos Informáticos.

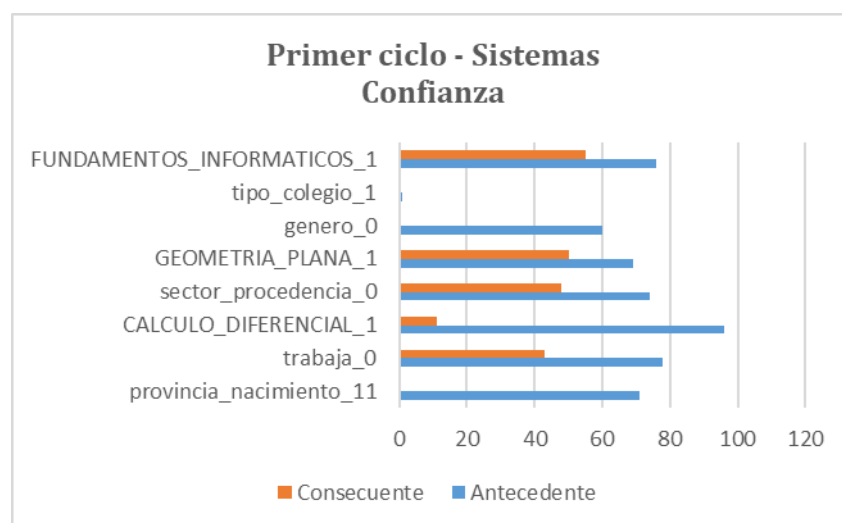


Figura 67. Confianza - Primer ciclo Sistemas

Lo mismo ocurre al seleccionar los resultados obtenidos por el indicador Lift, en ellos se muestra a estudiantes de género masculino, oriundos de la Provincia de Loja, no ejercer ninguna actividad laboral y aprobar las materias de Geometría Plana seguida de Cálculo Diferencial, una de las materias presentes en la malla vigente.

- **Segundo ciclo – Sistemas**

Para este modelo de han obtenido los siguientes resultados descritos en la Figura 68, donde se identifica que los estudiantes no deben contar con ninguna actividad laboral, pertenecer al género masculino, residir en el sector Urbano y haber culminado sus estudios secundarios en una Institución Fiscomisional

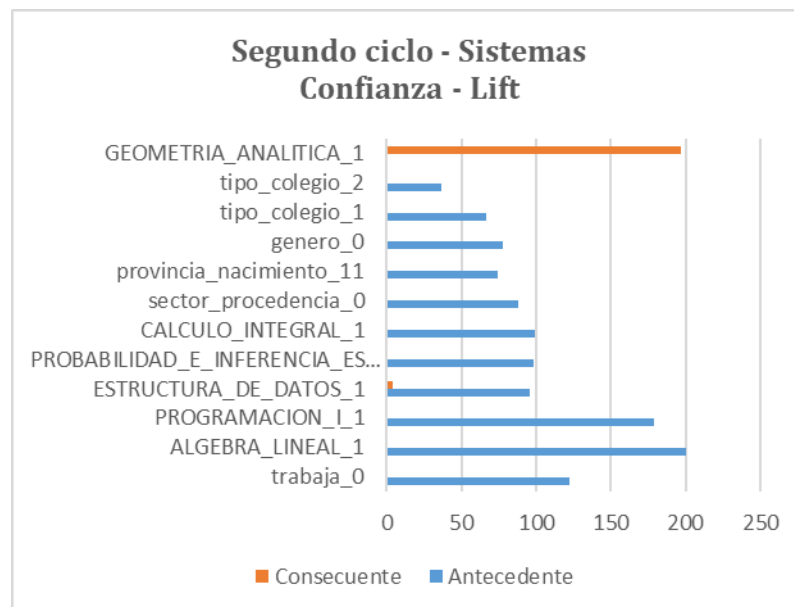


Figura 68. Confianza - Segundo ciclo Sistemas

- **Tercer ciclo – Sistemas**

Para el modelo del tercer ciclo (Ver Figura 69) y considerando como métrica a la confianza, se ha demostrado que la variable social que más sobresale son los estudiantes de género masculino, y a Colegios Fiscomisionales como Institución donde los estudiantes culminaron sus estudios secundarios esta característica como variable demográfica; las asignaturas que han tenido gran influencia dentro de este ciclo son: Estructura de datos II y Programación II. Estas características están

presentes en las reglas, dentro del antecedente como del consecuente, mismas que se relejan en las reglas obtenidas.

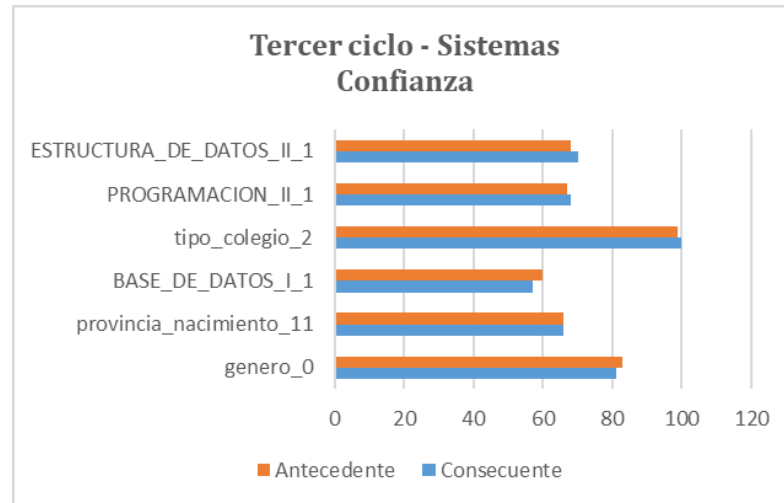


Figura 69. Confianza - Tercer ciclo Sistemas

Para el caso de Lift, se identifica una semejanza con las características obtenidas por la confianza, es decir, estudiantes que culminaron sus estudios secundarios en una Institución Fiscomisional, predomina el género masculino y las mismas asignaturas relacionadas a las matemáticas. En el caso de la Ingeniería del Software I, ya se muestra, pero no de manera representativa como las que anteceden.

- **Cuarto ciclo – Sistemas**

Para este ciclo, dentro de la Figura 70 se observa que, estudiantes que no ejercen ninguna actividad laboral, es una de las características más representativas en cuanto a variables sociales; otra de las características con valores elevados son para aquellos estudiantes que provienen del sector Urbano, esto como parte de las variables demográficas, finalmente para las características más representativas dentro de las variables académicas, se muestra a la materia de Metodología de la Programación en conjunto con Diseño y Gestión de Base de Datos como las asignaturas que más sobresalen.

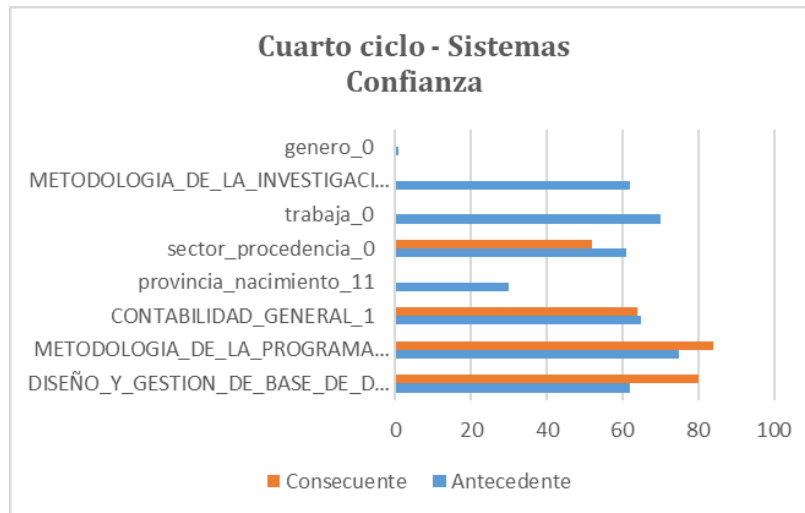


Figura 70. Confianza - Cuarto ciclo Sistemas

- **Quinto ciclo – Sistemas**

Las características que predominan en el quinto ciclo de acuerdo a la confianza tal y como se muestra en la Figura 71 se identifica que, estudiantes que no ejercen ninguna actividad laboral, es una de las características más representativas en cuanto a variables sociales; otra de las características con valores elevados son para aquellos estudiantes que provienen del sector Urbano, así como también, hace referencia a instituciones Fiscales y Fiscomisionales donde los estudiantes culminaron sus estudios secundarios, esto como parte de las variables demográficas, para este ciclo se identifica la presencia del género femenino como antecedente, sin embargo en este ciclo sigue predominado el género masculino en conjunto con las variables antes mencionadas, finalmente para las características más representativas dentro de las variables académicas, se muestra a la materia de Estadística Inferencial como la más predominante en ambos estados, aprobada y reprobada, mientras que, la aprobación de la asignatura de Programación Avanzada se da como consecuente, esto quiere decir que necesita de otros parámetros para ser aprobada con éxito.



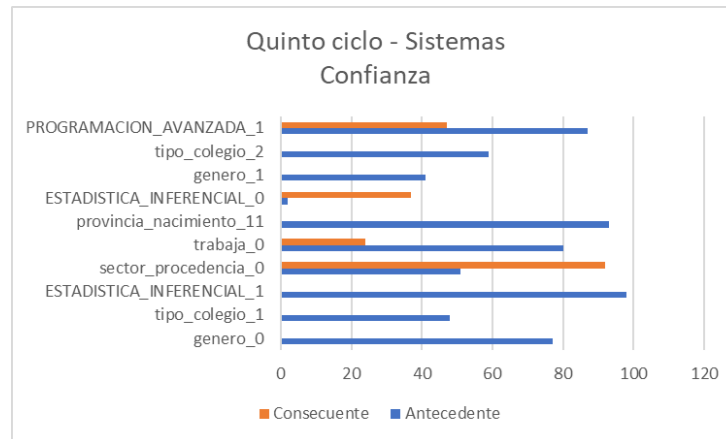


Figura 71. Confianza - Quinto ciclo Sistemas

- **Sexto ciclo – Sistemas**

Dentro del sexto ciclo, se encuentra como antecedente a la provincia de nacimiento y trabaja como las características más representativas, misma que hacer referencia a estudiantes que culminan sus estudios secundarios en una Institución Fiscomisional y no ejercen ninguna actividad laboral respectivamente, en el mismo rango, se visualiza a las materias tales como Análisis y Diseño de Sistemas y Lenguaje Ensamblador como las más representativas. Mientras que, como consecuente se tiene al sector de procedencia (Urbano) y a las materias de Simulación y Lenguaje Ensamblador (Ver Figura 72).

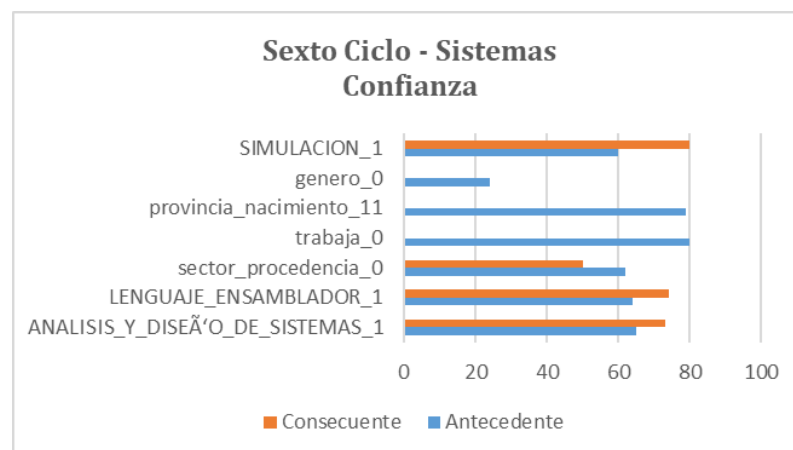


Figura 72. Confianza - Sexto ciclo Sistemas

Como se puede identificar en los resultados expuestos con anterioridad, las características coinciden con las premisas obtenidas al aplicar los algoritmos, en este caso en particular, mantiene a estudiantes de género

masculino y que no ejercen ninguna actividad laboral, en conjunto con el colegio donde culminaron sus estudios secundarios y, con ciertas asignaturas con estado de aprobadas, pero sin omitir la relación que guardan con cada una de estas características.

- **Séptimo ciclo – Sistemas**

Las características representativas para en séptimo ciclo correspondientes a variables sociales, académicas y demográficas en relación a la confianza con el antecedente tal y como se muestra en la Figura 73, es el sector de procedencia y trabaja, donde los estudiantes deben ser del sector Urbano y no contar con actividades laborales, además de las materias de Ecuaciones Diferenciales y Sistemas Operativos pertenecientes a la misma agrupación, por otro lado, como consecuente se tiene a sector de procedencia como la característica con más valor, esto seguida de las asignaturas de Diseño de Sistemas y Sistemas operativos.

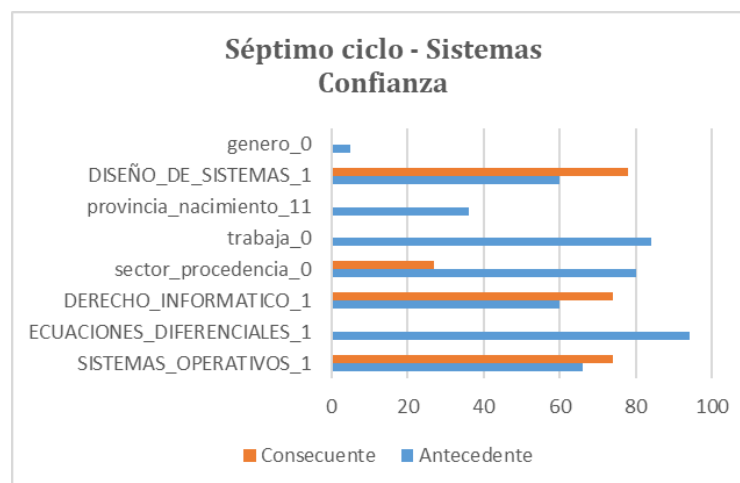


Figura 73. Confianza - Séptimo ciclo Sistemas

En cuanto al Lift, se demuestra al tipo de colegio, Institución Fiscomisional donde los estudiantes culminan sus estudios secundarios y trabaja, donde los alumnos no deben contar con ninguna actividad laboral; dentro de las asignaturas se observa a Sistemas operativos como la característica más predominante dentro de las características a considerar.

- **Octavo ciclo – Sistemas**

Con respecto al octavo ciclo, y en relación con la confianza se muestra dentro de la Figura 74, a trabaja, sector de procedencia y asignaturas como Análisis Numérico y Gestión de Redes, todas estas características como antecedente dentro de las premisas, mientras que sector de procedencia (Urbano) y Auditoria Informática en conjunto con Administración de Centro de Cómputo como cualidades predominantes del consecuente de una Regla.

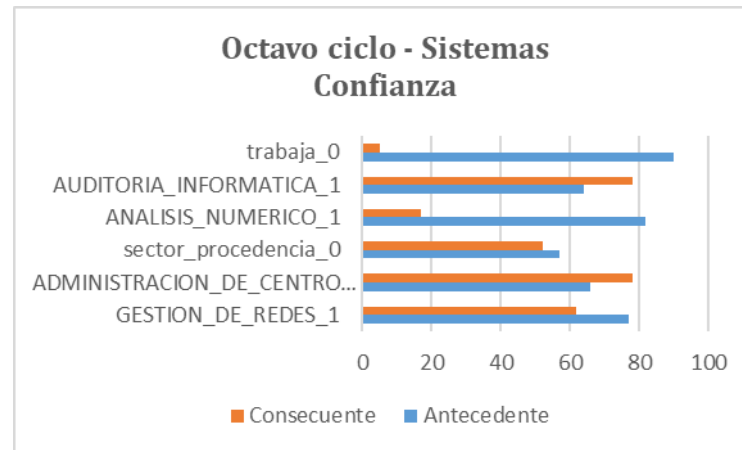


Figura 74. Confianza - Octavo ciclo Sistemas

- **Noveno ciclo – Sistemas**

Dentro del noveno ciclo (Ver Figura 75), poniendo en consideración la confianza, las características tales como no contar con ninguna actividad laboral, haber nacido en la provincia de Loja y aprobar a las asignaturas de Ingeniería del Software II y Modelamiento Matemático como las más representativas agrupadas en el antecedente, mientras que, pertenecer al sector Urbano y aprobar Ingeniería del Software II como atributos destacados en el Consecuente.

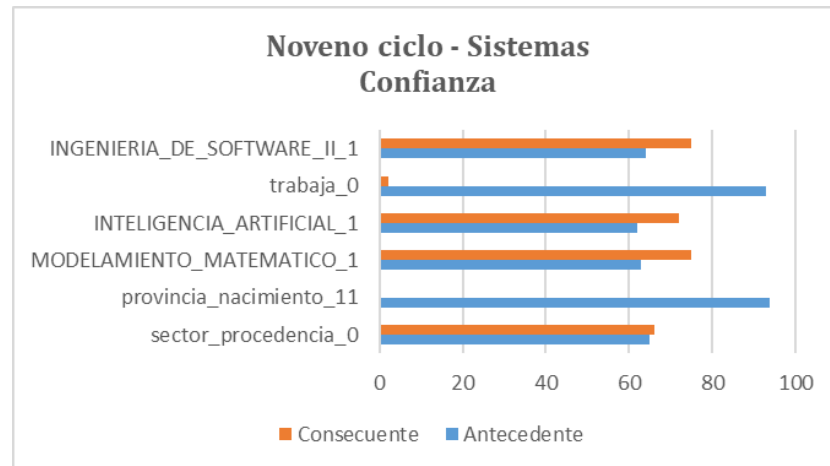


Figura 75. Confianza - Noveno ciclo Sistemas

- **Décimo ciclo – Sistemas**

Para el último ciclo de la Carrera de Ingeniería en Sistemas, se identifica dentro de la Figura 76, que las características de no contar con ninguna actividad laboral, ser parte del sector Urbano y aprobar las materias de Sistemas Expertos y Control Automatizado Asistido por Computadoras son las que predominan dentro de los atributos agrupados en los antecedentes y consecuentes

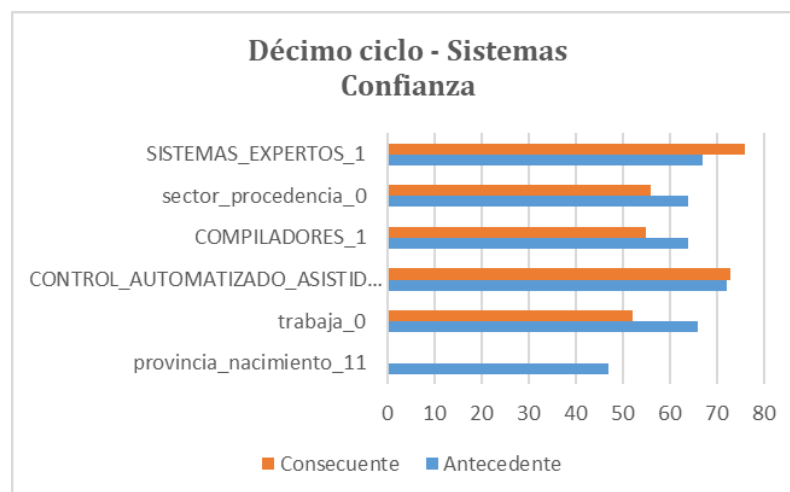


Figura 76. Confianza - Décimo ciclo Sistemas

Desde ciclos inferiores se demuestra mucho para aprobar a ciertas materias relacionadas con las matemáticas, los estudiantes deben ser de género masculino, pertenecer a un sector Urbano, haber culminado sus estudios secundarios en una Institución Fiscomisional y sobre todo se ha demostrado que, la característica de no contar con una actividad laboral presente en cada uno de



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

los ciclos es primordial en cada ciclo que cursan los estudiantes. Además, como se mencionó anteriormente, la presencia de mujeres haciendo referencia a ciclo inferiores es mínima, ya que al tratarse de asignaturas como Programación, Cálculo Diferencial, entre otras, no son aprobadas por las mismas, esto no significa que sus habilidades sean menores a la del género masculino, sin embargo, tras estudios se ha demostrado que las mujeres tiene un alto de aprendizaje en cuanto a la lectura, lo que puede explicar en este caso, la durabilidad de las mujeres que logran terminar una Carrera STEM, ya que, a partir del quinto ciclo ya se denota la presencia del género femenino, y las pocas que han llegado hasta esas instancias tienen más probabilidad de poder graduarse.

A continuación, se describen algunos de los resultados significativos que se han encontrado en cada uno de los ciclos de la Carrera de Ingeniería en Computación.

- **Primer ciclo – Computación**

Acerca de las características más representativas dentro de las variables sociales, académicas o demográficas y considerando como indicador a la confianza (Ver Figura 77), se tiene a género, donde existen más estudiantes de género masculino, a la característica sector de procedencia, misma que hace referencia al sector Urbano; tipo de colegio, en este caso las Instituciones Fiscomisionales son las que más predominan. Con referencia a las asignaturas se tiene como principal a la materia Introducción a las Ciencias de la Computación seguida de Comunicación Profesional, estas agrupadas en el antecedente. Mientras que, en el consecuente, el atributo trabaja, predomina como una de las características más relevantes dentro de esta agrupación seguida del tipo de colegio, haciendo referencia a una Institución Fiscomisional, finalmente a la aprobación de la materia de Comunicación Profesional.

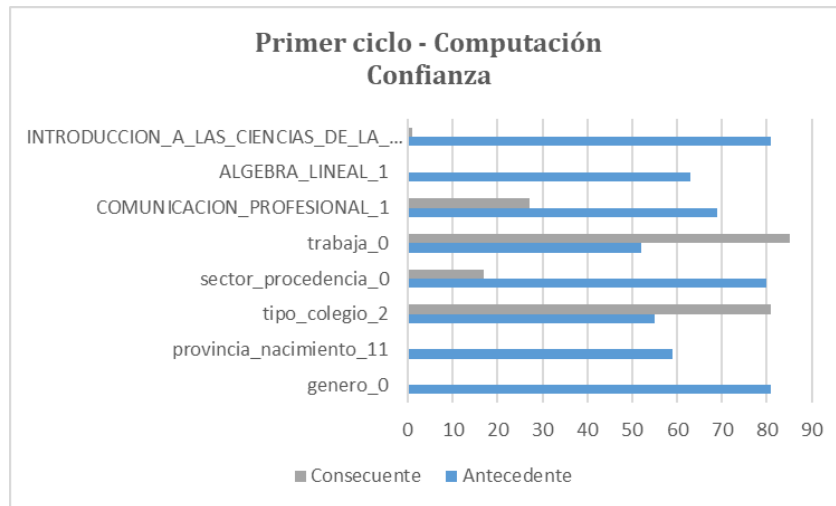


Figura 77. Confianza - Primer Ciclo Computación

- **Segundo ciclo – Computación**

Para el segundo ciclo, como se muestra en la Figura 78 se tiene a tipo colegio, la misma que hace referencia a Instituciones Fiscomisionales y a no ejercer ninguna actividad laboral, así como también la presencia de género masculino, con relación a las asignaturas se tiene a Programación Orientada a Objetos como la más predominante y como consecuente se tiene a no trabajar en conjunto con la aprobación de la materia de Cálculo diferencial como las características relevantes en esta sección.

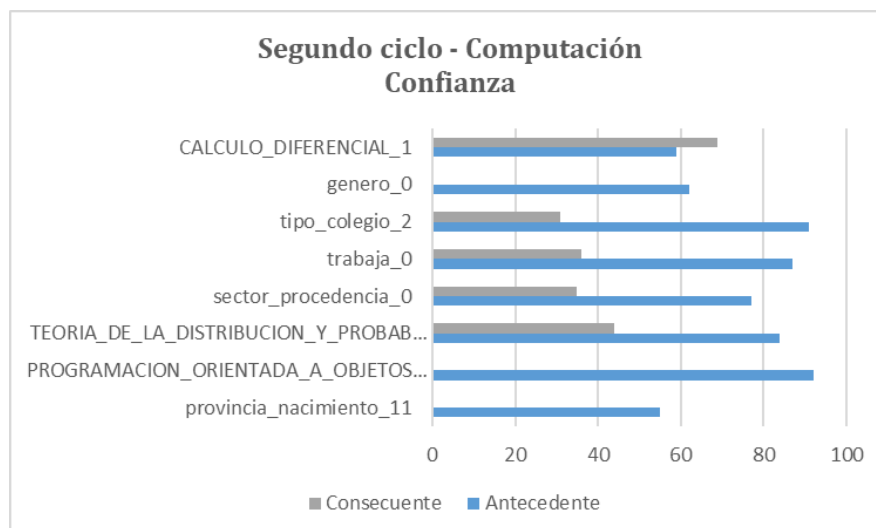


Figura 78. Confianza - Segundo Ciclo Computación

- **Tercer ciclo – Computación**

En cuanto a los resultados obtenidos con el indicador confianza y únicamente considerando a la confianza (Ver Figura 79), se identifica a género (masculino), trabajar (no contar con ninguna actividad laboral) y sector de procedencia (Urbano) como variables sociales y demográficas, mientras que, aprobar Estructura de Datos y Algoritmos Fundamentales y Estadística Analítica dentro de variables Académicas. En cambio, con el indicador Lift únicamente se obtienen resultados académicos, como son aprobar las materias de Cátedra Integradora: Proyecto de Vinculación “Centro de Asesoría Tecnológica y Cálculo Integral.

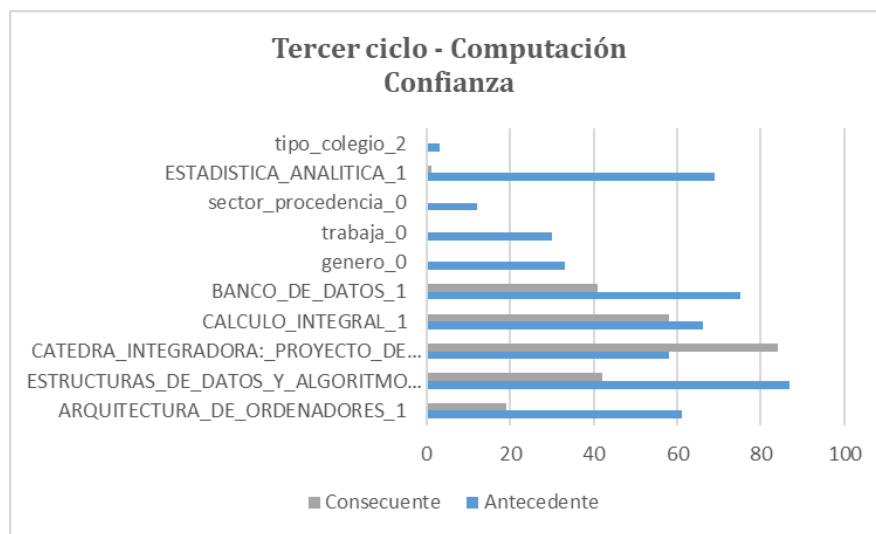


Figura 79. Confianza - Tercer Ciclo Computación

- **Cuarto ciclo – Computación**

Las características representativas para el último ciclo a tratar en este caso, correspondientes a variables sociales, académicas y demográficas en relación a la confianza y el antecedente es trabaja y género, donde los estudiantes de género masculino no deben realizar ninguna actividad laboral, así también aprobar a las materias de Estructura de Datos Avanzadas y Ecuaciones Diferenciales. Así mismo, conforme al consecuente tiene más representativas a las variables Académicas, donde las materias que sobresalen son Ingeniería de la Contaminación y Procesamiento de Transacciones, ambas aprobadas (Ver Figura 80).

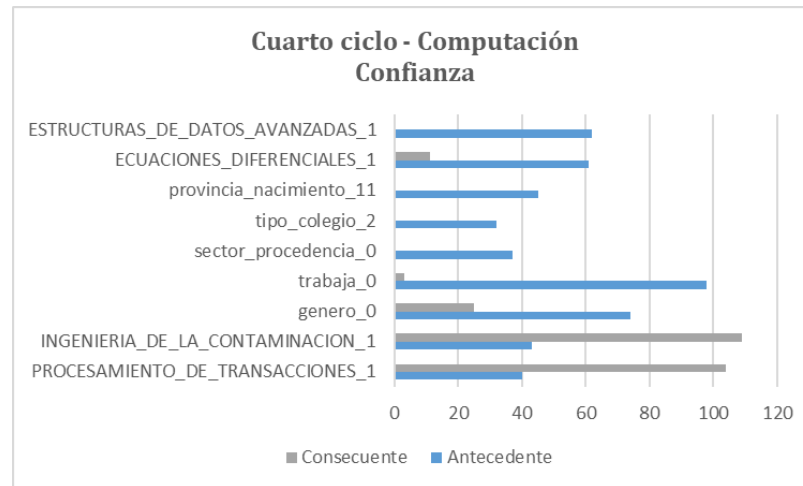


Figura 80. Confianza - Cuarto Ciclo Computación

## 6. Conclusión

- La aplicación de una metodología para el desarrollo del proceso de Minería de datos, presenta una gran importancia ya que si bien es cierto no asegura el proyecto perfecto, pero si emitirá las bases necesarias para que el proyecto se cumpla de manera eficiente y eficaz. Dentro del ámbito educativo es muy implementada para este tipo de inconvenientes, ya que proporciona una serie de fases con sus respectivas actividades que ayudarán a desarrollar el proyecto de la mejor manera y obteniendo buenos resultados.
- Con el algoritmo K-means se lograron obtener 5 clúster los cuales indican a nivel general el número de asignaturas perdidas a nivel de cada ciclo, con ello se da a conocer donde se encuentra el inconveniente, considerando que dentro de los clústers asignados no se toma en consideración al género femenino, lo que indica un bajo número de las mismas dentro de Carreras técnicas.
- Dentro del algoritmo A priori, las reglas generadas se basan en asignaturas con más iteraciones repetidas, es decir, hay asignaturas pertenecientes a una malla que no constan en la otra, tal es el caso Laboratorio de Física I, Laboratorio de Física II, Teoría de la programación, las demás como es el caso de Proyectos Informáticos se debe a que no muchos estudiantes han cursado la asignatura, por ende, se mantiene con ese valor.



#### Anexo 4. Ingeniería en Sistemas – Primer ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo para el **primer ciclo de la Carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

#### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y bajo decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

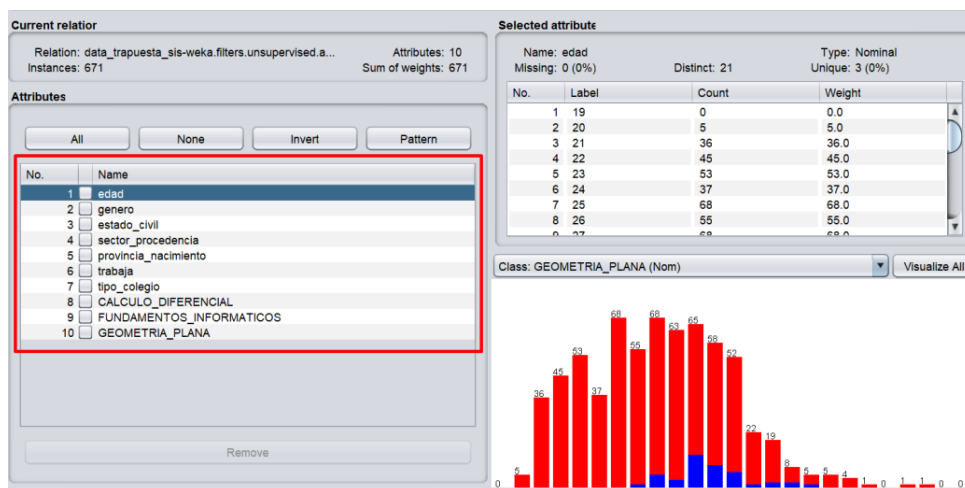


Figura 81. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

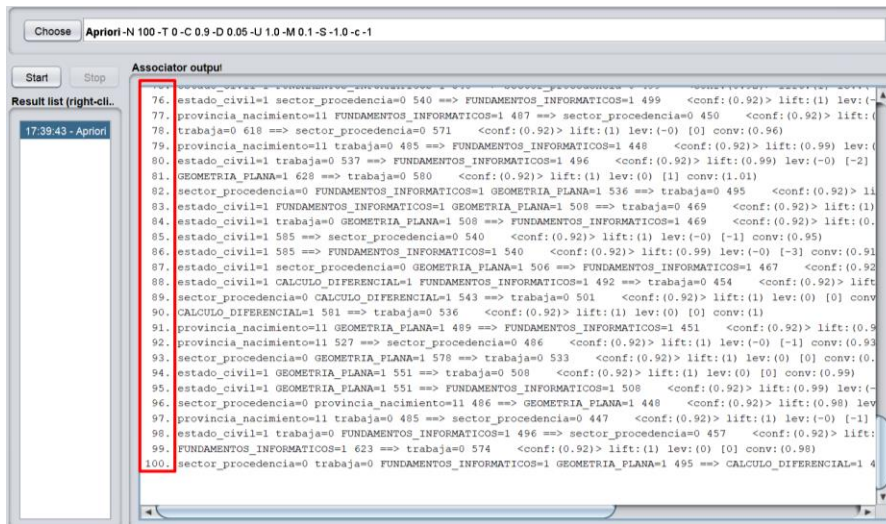


Figura 82. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza mayor a 0.9 y lift mayor a 1 respectivamente, se obtiene un total de 16 reglas (Ver Figura 3).

REGLAS	CONFIANZ.	LIFT		
1. trabaja=0 CALCULO_DIFERENCIAL=1 536 ==> FUNDAMENTOS_INFORMATICOS=1 524	conf:(0.98)	lift:(1.05)	lev:(0.04)	[26]
2. estado_civil=1 trabaja=0 CALCULO_DIFERENCIAL=1 465 ==> FUNDAMENTOS_INFORMATICOS=1 454	conf:(0.98)	lift:(1.05)	lev:(0.03)	[22]
3. trabaja=0 CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 503 ==> FUNDAMENTOS_INFORMATICOS=1 491	conf:(0.98)	lift:(1.05)	lev:(0.04)	[23]
4. sector_procedencia=0 trabaja=0 CALCULO_DIFERENCIAL=1 501 ==> FUNDAMENTOS_INFORMATICOS=1 489	conf:(0.98)	lift:(1.05)	lev:(0.04)	[23]
5. CALCULO_DIFERENCIAL=1 581 ==> FUNDAMENTOS_INFORMATICOS=1 567	conf:(0.98)	lift:(1.05)	lev:(0.04)	[27]
6. provincia_nacimiento=11 CALCULO_DIFERENCIAL=1 450 ==> FUNDAMENTOS_INFORMATICOS=1 439	conf:(0.98)	lift:(1.05)	lev:(0.03)	[21]
7. sector_procedencia=0 trabaja=0 CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 468 ==> FUNDAMENTOS_INFORMATICOS=1 456	conf:(0.97)	lift:(1.05)	lev:(0.03)	[21]
8. estado_civil=1 CALCULO_DIFERENCIAL=1 505 ==> FUNDAMENTOS_INFORMATICOS=1 492	conf:(0.97)	lift:(1.05)	lev:(0.03)	[23]
9. sector_procedencia=0 CALCULO_DIFERENCIAL=1 543 ==> FUNDAMENTOS_INFORMATICOS=1 529	conf:(0.97)	lift:(1.05)	lev:(0.04)	[24]
10. CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 543 ==> FUNDAMENTOS_INFORMATICOS=1 529	conf:(0.97)	lift:(1.05)	lev:(0.04)	[24]
11. estado_civil=1 CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 475 ==> FUNDAMENTOS_INFORMATICOS=1 462	conf:(0.97)	lift:(1.05)	lev:(0.03)	[20]
12. estado_civil=1 sector_procedencia=0 CALCULO_DIFERENCIAL=1 471 ==> FUNDAMENTOS_INFORMATICOS=1 458	conf:(0.97)	lift:(1.05)	lev:(0.03)	[20]
13. sector_procedencia=0 CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 505 ==> FUNDAMENTOS_INFORMATICOS=1 491	conf:(0.97)	lift:(1.05)	lev:(0.03)	[22]
14. estado_civil=1 trabaja=0 CALCULO_DIFERENCIAL=1 465 ==> GEOMETRIA_PLANA=1 440	conf:(0.95)	lift:(1.01)	lev:(0.01)	[4]
15. estado_civil=1 trabaja=0 537 ==> GEOMETRIA_PLANA=1 508	conf:(0.95)	lift:(1.01)	lev:(0.01)	[5]
16. estado_civil=1 trabaja=0 FUNDAMENTOS_INFORMATICOS=1 496 ==> GEOMETRIA_PLANA=1 469	conf:(0.95)	lift:(1.01)	lev:(0.01)	[4]

Figura 83. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante no trabaja y aprueba la asignatura de Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.



- 2) Si un estudiante con estado civil soltero, no trabaja y aprueba Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.
- 3) Si un estudiante no trabaja y aprueba Cálculo Diferencial y Geometría Plana, este aprobará Fundamentos Informáticos.
- 4) Si un estudiante es del sector Urbano, no trabaja, aprueba Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.
- 5) Si un estudiante aprueba Cálculo Diferencial, también aprobará Fundamentos Informáticos.
- 6) Si un estudiante perteneciente a la Provincia de Loja y aprueba Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.
- 7) Un estudiante que pertenezca al sector Urbano, no trabaje, apruebe Cálculo Diferencial y Geometría Plana, entonces aprobará Fundamentos Informáticos.
- 8) Estudiante que apruebe Cálculo Diferencial y que sea soltero, aprobará Fundamentos Informáticos.
- 9) Si un estudiante con estado civil soltero, aprueba la asignatura de Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.
- 10) Si un estudiante aprueba Cálculo Diferencial y Geometría Plana, entonces aprobará Fundamentos Informáticos.
- 11) Si un estudiante con estado civil soltero, aprueba Cálculo Diferencial y Geometría Plana, entonces aprobará Fundamentos Informáticos.
- 12) Si un estudiante es soltero, procedente del sector Urbano y aprueba Cálculo Diferencial, entonces aprobará Fundamentos Informáticos.
- 13) Si un estudiante procedente del sector Urbano y aprueba Cálculo Diferencial y Geometría Plana, también aprobará Fundamentos Informáticos.
- 14) Si un estudiante soltero, no trabaja y aprueba Cálculo Diferencial entonces aprobará Geometría Plana.
- 15) Si un estudiante es soltero y no trabaja, aprobará Geometría Plana.
- 16) Si un estudiante soltero, no trabaja y aprueba Fundamentos Informáticos, también aprobará Geometría Plana.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado con valore de 1.1 (Ver Figura 4) con lo que se obtienen los siguientes resultados:

1	REGLAS	B	C	D	E
16	15. trabaja=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 224 ==> sector_procedencia=0 FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 207	con 0.92	lift: (1.16)	lev: (0.04) [28]	conv: (2.5)
21	20. sector_procedencia=0 trabaja=0 tipo_colegio=2 FUNDAMENTOS_INFORMATICOS=1 222 ==> CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 207	con 0.93	lift: (1.15)	lev: (0.04) [27]	conv: (2.6)
26	25. estado_civil=1 tipo_colegio=2 FUNDAMENTOS_INFORMATICOS=1 225 ==> CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 208 conf:	0.92	lift: (1.14)	lev: (0.04) [25]	conv: (2.38)
28	27. sector_procedencia=0 tipo_colegio=2 FUNDAMENTOS_INFORMATICOS=1 238 ==> CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 220 conf:	0.92	lift: (1.14)	lev: (0.04) [27]	conv: (2.38)
30	29. trabaja=0 tipo_colegio=2 FUNDAMENTOS_INFORMATICOS=1 235 ==> CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 217 conf:	0.92	lift: (1.14)	lev: (0.04) [26]	conv: (2.36)
32	31. tipo_colegio=2 CALCULO_DIFERENCIAL=1 242 ==> sector_procedencia=0 FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 220 conf:	0.91	lift: (1.14)	lev: (0.04) [26]	conv: (2.12)
42	41. tipo_colegio=2 FUNDAMENTOS_INFORMATICOS=1 253 ==> CALCULO_DIFERENCIAL=1 GEOMETRIA_PLANA=1 232 conf:	0.92	lift: (1.13)	lev: (0.04) [27]	conv: (2.15)
52	51. sector_procedencia=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 230 ==> trabaja=0 FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 207	con 0.9	lift: (1.12)	lev: (0.03) [22]	conv: (1.9)
58	57. tipo_colegio=2 CALCULO_DIFERENCIAL=1 242 ==> trabaja=0 FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 217 conf:	0.9	lift: (1.12)	lev: (0.03) [22]	conv: (1.84)
62	61. trabaja=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 224 ==> FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 217 conf:	0.97	lift: (1.12)	lev: (0.03) [22]	conv: (3.7)
64	63. sector_procedencia=0 trabaja=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 214 ==> FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 207	con 0.97	lift: (1.12)	lev: (0.03) [21]	conv: (3.55)
86	85. tipo_colegio=2 CALCULO_DIFERENCIAL=1 242 ==> FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 232 conf:	0.96	lift: (1.11)	lev: (0.03) [22]	conv: (2.92)
96	95. sector_procedencia=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 230 ==> FUNDAMENTOS_INFORMATICOS=1 GEOMETRIA_PLANA=1 220 conf:	0.96	lift: (1.1)	lev: (0.03) [20]	conv: (2.7)

Figura 84. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante no trabaja, ha culminado sus estudios en un colegio Fiscomisional, aprueba Cálculo Diferencial, es muy probable que sea del sector Urbano y que aprobará Fundamentos Informáticos y Geometría Plana.
- 2) Estudiantes que procedan del sector Urbano, no se encuentren laborando y aprueben Fundamentos Informáticos, probablemente aprobarán Cálculo Diferencial y Geometría Plana.
- 3) Estudiantes que sean solteros y que procedan de un Colegio Fiscomisional, aprueban Fundamentos Informáticos, también es muy probable que aprueben Cálculo Diferencial y Geometría Plana.
- 4) Estudiantes que sean solteros y que procedan de un Colegio Fiscomisional, aprueban Fundamentos Informáticos, también es muy probable que aprueben Cálculo Diferencial y Geometría Plana.
- 5) Estudiantes que no trabajen y que procedan de un Colegio Fiscomisional, aprueban Fundamentos Informáticos, también es muy probable que aprueben Cálculo Diferencial y Geometría Plana.
- 6) Si un estudiante realizó sus estudios en un colegio Fiscomisional y aprueba Cálculo Diferencial es muy probable que dicho estudiante pertenezca al sector Urbano y que a su vez apruebe Fundamentos Informáticos y Geometría Plana.
- 7) Estudiante graduado de una institución educativa Fiscomisional y apruebe Fundamentos Informáticos también es probable que apruebe Cálculo Diferencial y Geometría Plana.
- 8) Si un estudiante pertenece al sector Urbano y es graduado de un colegio Fiscomisional, aprueban Cálculo Diferencial es muy posible que este no trabaje y también apruebe Fundamentos Informáticos y Geometría Plana.

- 9) Si un estudiante es graduado de un colegio Fiscomisional y aprueba Cálculo Diferencial es muy posible que este no trabaje y también apruebe Fundamentos Informáticos y Geometría Plana.
- 10) Si un estudiante no trabaja, es graduado de un colegio Fiscomisional y aprueba Cálculo Diferencial es muy posible que este apruebe Fundamentos Informáticos y Geometría Plana.
- 11) Si un estudiante pertenece al sector Urbano, no trabaja, es graduado de un colegio Fiscomisional y aprueba Cálculo Diferencial es muy posible que también apruebe Fundamentos Informáticos y Geometría Plana.
- 12) Si un estudiante es graduado de un colegio Fiscomisional y aprueba Cálculo Diferencial es muy probable que también apruebe Fundamentos Informáticos y Geometría Plana.
- 13) Si un estudiante pertenece al sector Urbano, es graduado de un colegio Fiscomisional y aprueba Cálculo Diferencial posiblemente también apruebe Fundamentos Informáticos y Geometría Plana.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano, considerando que hay muy pocos estudiantes del sector Rural, están muy presentes dentro de las premisas obtenidas. De la misma manera, se denota ausencia del género femenino aun tomando en cuenta una confianza del 88%, esto debido a que el número de mujeres presentes en este tipo de Ingenierías es mínimo.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas son muy similares a los resultados obtenidos en la herramienta weka. Se mantiene la relación entre el género, sector de procedencia, el tipo de colegio, si trabaja o no, y si este aprueba materias vinculadas a otras. Esto se ve reflejado en la Figura 5.



## AssociationRules

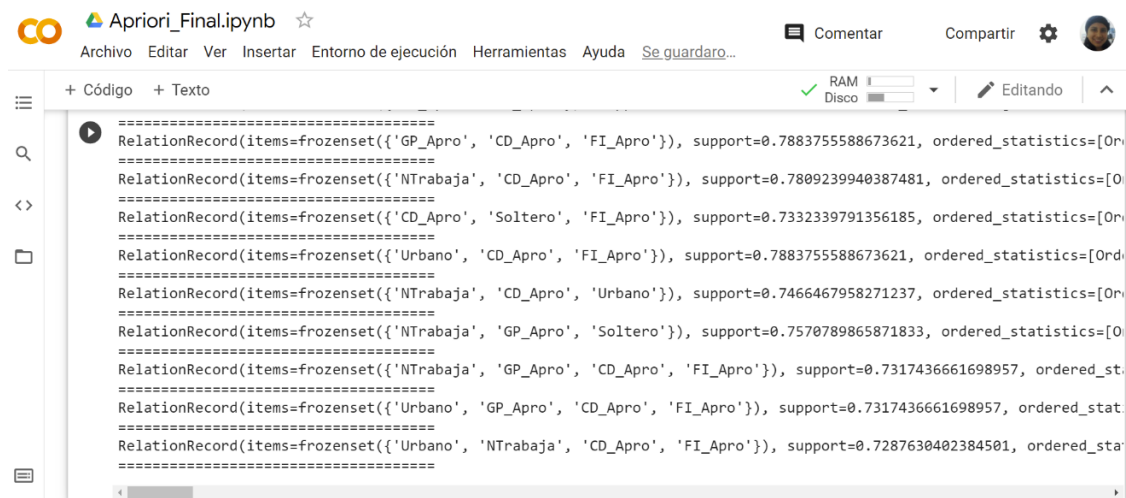
Association Rules

```
[GEOMETRIA PLANA, genero] --> [estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.741)
[FUNDAMENTOS_INFORMATICOS] --> [GEOMETRIA_PLANA, estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.742)
[CALCULO_DIFERENCIAL, tipo_colegio, edad] --> [estado_civil] (confidence: 0.743)
[GEOMETRIA_PLANA, tipo_colegio, edad] --> [estado_civil] (confidence: 0.743)
[provincia_nacimiento] --> [GEOMETRIA_PLANA, estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.744)
[tipo_colegio, edad] --> [estado_civil] (confidence: 0.744)
[CALCULO_DIFERENCIAL, genero] --> [GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, estado_civil] (confidence: 0.747)
[tipo_colegio] --> [FUNDAMENTOS_INFORMATICOS, estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.748)
[GEOMETRIA_PLANA, tipo_colegio] --> [FUNDAMENTOS_INFORMATICOS, estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.753)
[tipo_colegio] --> [GEOMETRIA_PLANA, estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.754)
[genero] --> [GEOMETRIA_PLANA, estado_civil] (confidence: 0.754)
[FUNDAMENTOS_INFORMATICOS, genero] --> [GEOMETRIA_PLANA, estado_civil] (confidence: 0.754)
[GEOMETRIA_PLANA] --> [estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.756)
[estado_civil, edad] --> [GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, CALCULO_DIFERENCIAL] (confidence: 0.756)
[GEOMETRIA_PLANA, CALCULO_DIFERENCIAL, edad] --> [FUNDAMENTOS_INFORMATICOS, estado_civil] (confidence: 0.756)
[CALCULO_DIFERENCIAL, genero] --> [GEOMETRIA_PLANA, estado_civil] (confidence: 0.759)
[FUNDAMENTOS_INFORMATICOS, genero] --> [estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.760)
[FUNDAMENTOS_INFORMATICOS, CALCULO_DIFERENCIAL, genero] --> [GEOMETRIA_PLANA, estado_civil] (confidence: 0.761)
[CALCULO_DIFERENCIAL, edad] --> [FUNDAMENTOS_INFORMATICOS, estado_civil] (confidence: 0.761)
[GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, edad] --> [estado_civil] (confidence: 0.763)
[GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, CALCULO_DIFERENCIAL, edad] --> [estado_civil] (confidence: 0.763)
[edad] --> [GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, CALCULO_DIFERENCIAL] (confidence: 0.763)
[FUNDAMENTOS_INFORMATICOS, edad] --> [estado_civil] (confidence: 0.765)
[estado_civil, edad] --> [GEOMETRIA_PLANA, CALCULO_DIFERENCIAL] (confidence: 0.76)
[GEOMETRIA_PLANA, CALCULO_DIFERENCIAL, edad] --> [estado_civil] (confidence: 0.765)
[GEOMETRIA_PLANA, FUNDAMENTOS_INFORMATICOS, genero] --> [estado_civil, CALCULO_DIFERENCIAL] (confidence: 0.766)
[FUNDAMENTOS_INFORMATICOS, CALCULO_DIFERENCIAL, edad] --> [estado_civil] (confidence: 0.767)
```

Figura 85. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está ejerciendo alguna actividad laboral, si su estado civil es soltero, si este pertenece al sector Urbano y la aprobación de algunas materias relacionadas de forma paralela.



```
RelationRecord(items=frozenset({'GP_Apro', 'CD_Apro', 'FI_Apro'}), support=0.7883755588673621, ordered_statistics=[0.7883755588673621, 0.7883755588673621, 0.7883755588673621])
RelationRecord(items=frozenset({'NTrabaja', 'CD_Apro', 'FI_Apro'}), support=0.7809239940387481, ordered_statistics=[0.7809239940387481, 0.7809239940387481, 0.7809239940387481])
RelationRecord(items=frozenset({'CD_Apro', 'Soltero', 'FI_Apro'}), support=0.7332339791356185, ordered_statistics=[0.7332339791356185, 0.7332339791356185, 0.7332339791356185])
RelationRecord(items=frozenset({'Urbano', 'CD_Apro', 'FI_Apro'}), support=0.7883755588673621, ordered_statistics=[0.7883755588673621, 0.7883755588673621, 0.7883755588673621])
RelationRecord(items=frozenset({'NTrabaja', 'CD_Apro', 'Urbano'}), support=0.7466467958271237, ordered_statistics=[0.7466467958271237, 0.7466467958271237, 0.7466467958271237])
RelationRecord(items=frozenset({'NTrabaja', 'GP_Apro', 'Soltero'}), support=0.7570789865871833, ordered_statistics=[0.7570789865871833, 0.7570789865871833, 0.7570789865871833])
RelationRecord(items=frozenset({'NTrabaja', 'GP_Apro', 'CD_Apro', 'FI_Apro'}), support=0.7317436661698957, ordered_statistics=[0.7317436661698957, 0.7317436661698957, 0.7317436661698957, 0.7317436661698957])
RelationRecord(items=frozenset({'Urbano', 'GP_Apro', 'CD_Apro', 'FI_Apro'}), support=0.7317436661698957, ordered_statistics=[0.7317436661698957, 0.7317436661698957, 0.7317436661698957, 0.7317436661698957])
RelationRecord(items=frozenset({'Urbano', 'NTrabaja', 'CD_Apro', 'FI_Apro'}), support=0.728763042384501, ordered_statistics=[0.728763042384501, 0.728763042384501, 0.728763042384501, 0.728763042384501])
```

Figura 86. Reglas de asociación obtenidas por Python

## Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster (esto a decisión del investigador), cada uno de ellos con las instancias correspondientes, Ver Figura 7.

```

Clustered Instances

0      369 ( 55%)
1       10 (  1%)
2       42 (  6%)
3      235 ( 35%)
4        15 (  2%)
  
```

Figura 87. Clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 3. La Figura 8, indica las características que componen a cada clúster, a continuación, la descripción de cada uno de ellos:

```

Final cluster centroids:

Attribute          Full Data          Cluster#
                   (671.0)           0           1           2           3           4
                   (369.0)       (10.0)       (42.0)       (235.0)       (15.0)
=====
edad                25                25           25           29           22           21
genero              0                 0            1            0            0            0
estado_civil        1                 1            1            1            1            1
sector_procedencia  0                 0            0            0            0            0
provincia_nacimiento 11                11           11           11           11           7
trabaja             0                 0            1            0            0            0
tipo_colegio        1                 1            2            1            2            2
CALCULO_DIFERENCIAL 1                  1            1            1            1            1
FUNDAMENTOS_INFORMATICOS 1                 1            0            1            1            1
GEOMETRIA_PLANA    1                 1            1            0            1            1
  
```

Figura 88. Clústeres obtenidos por Weka.

**Clúster\_0 (55%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, es decir, Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana.



**Clúster\_1 (1%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género femenino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que se encuentren realizando cualquier actividad laboral junto a sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán únicamente dos materias relacionadas a las matemáticas correspondientes al primer ciclo, es decir, Cálculo Diferencial y Geometría Plana, mientras que, la asignatura de Fundamentos Informáticos no la aprueban.

**Clúster\_2 (6%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán dos de tres materias relacionadas a las matemáticas correspondientes al primer ciclo, es decir, Cálculo Diferencial y Fundamentos Informáticos, sin embargo la materia de Geometría Plana es reprobada.

**Clúster\_3 (35%):** Estudiantes que tengan 22 años de edad, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan culminado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, es decir, Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana.

**Clúster\_4 (2%):** Estudiantes con una edad de 21 años, de género masculino, de estado civil soltero, provenientes de la provincia de El Oro, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, es decir, Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana.

**Resultados obtenidos por Algoritmo K - means – RapidMiner**



La Figura 9 muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clúster.

Nominal values

Index	Nominal value	Absolute count	Fraction
1	cluster_0	469	0.699
2	cluster_1	69	0.103
3	cluster_2	46	0.069
4	cluster_3	46	0.069
5	cluster_4	41	0.061

Figura 89. Clúster con total de instancias

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 1 con más instancias agrupadas.

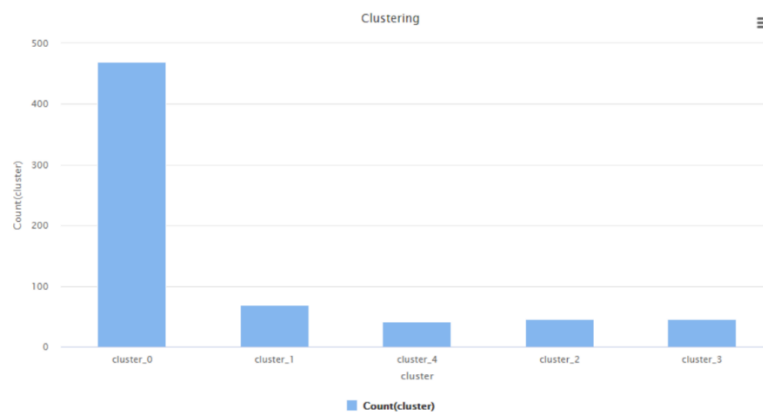


Figura 90. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

### k-Means - Centroid Table

Cluster	CALCULO ...	edad	estado_civil	FUNDAME...	genero	GEOMETRI...	provincia_...	sector_pr...	tipo_colegio	trabaja
Cluster 0	0.826	27.000	0.891	1	0.174	1	11.522	1	1.500	0.065
Cluster 1	0.946	25.507	1	1	0.202	1	11.529	-0	2.076	0.067
Cluster 2	0.889	27.242	1	1	0.316	1	11.398	-0	1.004	0.090
Cluster 3	0.292	25.750	0.938	-0.000	0.167	0.958	11.667	0.083	1.604	0.083
Cluster 4	0.918	30.291	0.291	1	0.373	0.627	11.509	0	1.445	0.082

Figura 91. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana, todas pertenecientes al primer ciclo, si tiene una edad de 27 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana, todas pertenecientes al primer ciclo, si tiene una edad de 25 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana, todas pertenecientes al primer ciclo, si tiene una edad de 27 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 3:** Estudiante aprueba únicamente la asignatura de Geometría plana y reprueba Cálculo Diferencial y Fundamentos Informáticos, todas pertenecientes al primer ciclo, si tiene una edad de 25 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Fundamentos Informáticos y Geometría Plana, todas pertenecientes al primer ciclo si tiene una edad de 30 años, es casado, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas y a su familia, es decir, no ejerce una actividad laboral.

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 12, se han obtenido los clústers con las variables **Tipo de colegio** y **Género**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.

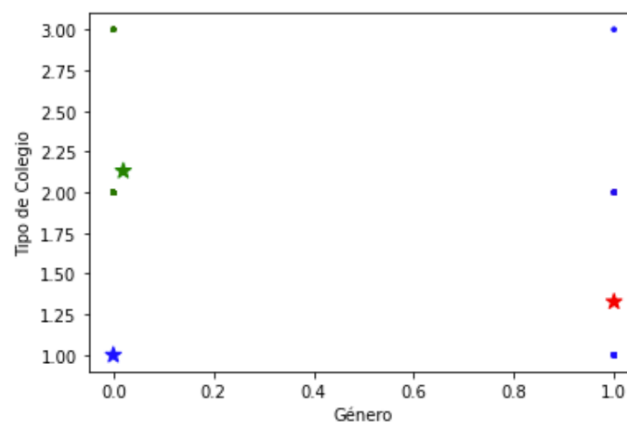


Figura 92. Relación Género - Tipo de Colegio

Por otra parte, la Figura 13 muestra la relación entre la variable Cálculo Diferencial y Género, donde se identifica que las mujeres no pierden dicha asignatura, mientras que dentro de los clústers restantes se identifica la reprobación y aprobación de la misma por parte del género masculino.

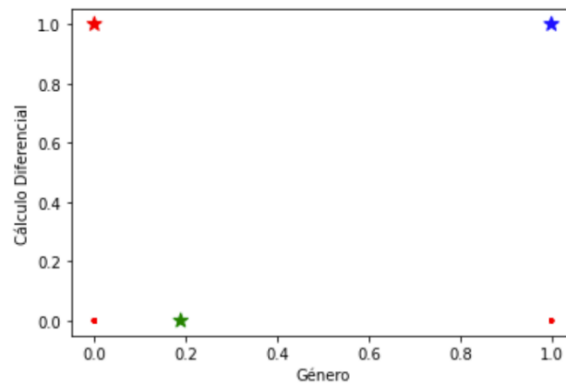


Figura 93. Relación Cálculo Diferencial - Género.

La figura 14, hace referencia a las variables **Edad** y **Fundamentos Informáticos**, en donde se identifica que se aprueba a dicha asignatura si los estudiantes se encuentran dentro de las edades de 23, 27 y 31 años de edad.

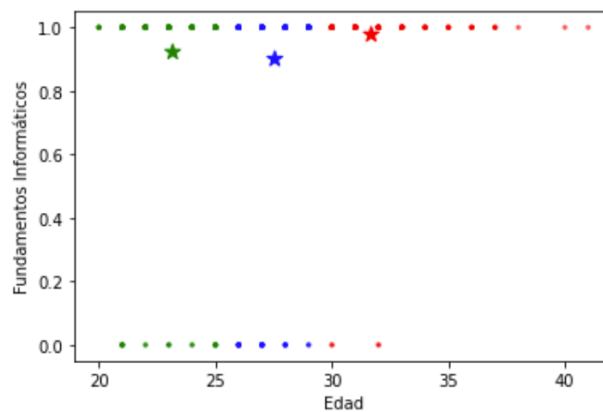


Figura 94. Relación Edad - Fundamentos Informáticos

Por otra parte, se ha realizado un análisis entre las variables **sector de procedencia** y **Fundamentos Informáticos**, se identifica a estudiantes que pertenecen al sector Rural reprobaban esta asignatura, mientras que otra lo aprueba, tomando en cuenta que pertenece al mismo sector. Se muestra también la aprobación de la materia por estudiantes del sector Urbano.

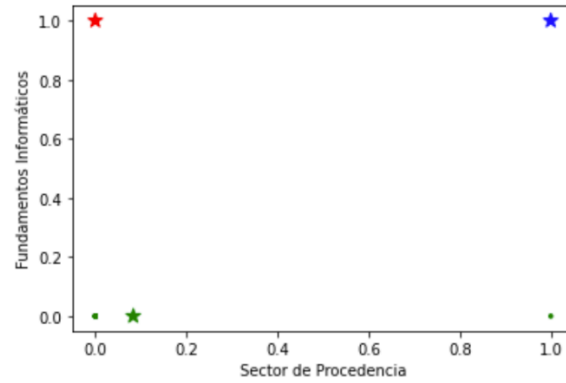


Figura 95. Relación Sector de procedencia - Fundamentos Informáticos

Se puede realizar cualquier tipo de relación entre variables y en base a estas se pueden obtener otros resultados en particular, esto se puede definir un poco más si se considera la correlación que existen entre ellas.

## Anexo 5. Ingeniería en Sistemas – Segundo ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo para el **segundo ciclo de la Carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda, clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

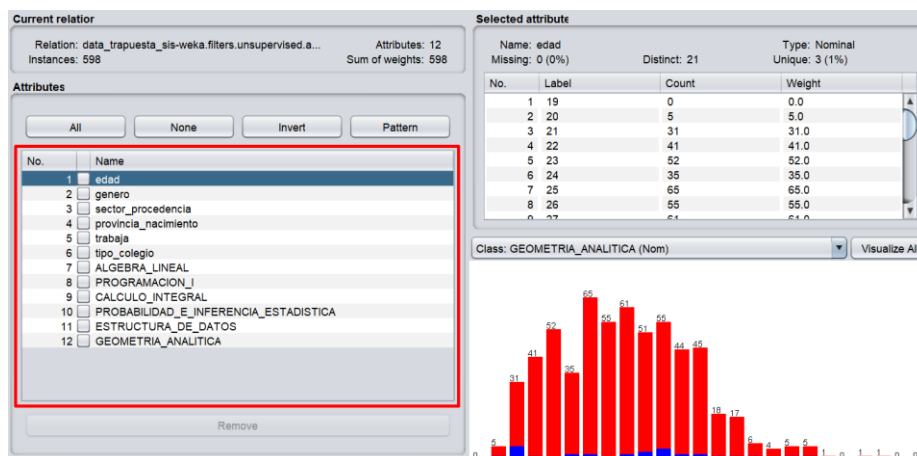


Figura 96. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_E
177. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTUR
178. sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEG
179. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_E
180. provincia_nacimiento=11 trabaja=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBA
181. provincia_nacimiento=11 trabaja=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 ESTRU
182. genero=0 tipo_colegio=2 ALGEBRA_LINEAL=1 164 ==> GEOMETRIA_ANALITICA=1 164 <conf:(1)> lift:(1.05) lev:(0
183. sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEG
184. provincia_nacimiento=11 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 163 ==> GEOMETRIA_ANALITICA=1 163 <con
185. sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEG
186. provincia_nacimiento=11 trabaja=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBA
187. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 162 ==>
188. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 ESTRUCTU
189. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 160 ==> ESTRUCTURA_DE_DATOS=1
190. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 160 ==> GEOMETRIA_ANALITICA=1
191. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 GEOMETRIA_ANALITICA=1 160 ==>
192. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 160 ==>
193. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 160 ==> ESTRUCTURA_DE_DATOS=1
194. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBABIL
195. genero=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 159 ==> GEOMETRIA_ANALITIC
196. genero=0 tipo_colegio=2 ALGEBRA_LINEAL=1 ESTRUCTURA_DE_DATOS=1 159 ==> GEOMETRIA_ANALITICA=1 159 <conf:(
197. provincia_nacimiento=11 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 1
198. sector_procedencia=0 trabaja=0 tipo_colegio=2 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBABIL
199. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTA
200. genero=0 tipo_colegio=1 ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTA
  
```

Figura 97. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza 1, equivalente al 100% de confiabilidad en la premisa, y su lift mayor a 1.05, se obtiene un total de 13 reglas (Ver Figura 3).

REGLAS	CONF.	LIFT
1. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 471 ==> GEOMETRIA_ANALITICA=1 470	conf:(1)	lift:(1.05)
2. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 ESTRUCTURA_DE_DATOS=1 467 ==> GEOMETRIA_ANALITICA=1 466	conf:(1)	lift:(1.05)
3. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 465 ==> GEOMETRIA_ANALITICA=1 464	conf:(1)	lift:(1.05)
4. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1 461 ==> GEOMETRIA_ANALITICA=1 460	conf:(1)	lift:(1.05)
5. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 453 ==> GEOMETRIA_ANALITICA=1 452	conf:(1)	lift:(1.05)
6. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 450 ==> GEOMETRIA_ANALITICA=1 449	conf:(1)	lift:(1.05)
7. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 478 ==> GEOMETRIA_ANALITICA=1 476	conf:(1)	lift:(1.05)
8. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 471 ==> GEOMETRIA_ANALITICA=1 469	conf:(1)	lift:(1.05)
9. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 467 ==> GEOMETRIA_ANALITICA=1 465	conf:(1)	lift:(1.05)
10. trabaja=0 ALGEBRA_LINEAL=1 465 ==> GEOMETRIA_ANALITICA=1 463	conf:(1)	lift:(1.05)
11. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1 461 ==> GEOMETRIA_ANALITICA=1 459	conf:(1)	lift:(1.05)
12. trabaja=0 ALGEBRA_LINEAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 454 ==> GEOMETRIA_ANALITICA=1 452	conf:(1)	lift:(1.05)
13. trabaja=0 ALGEBRA_LINEAL=1 ESTRUCTURA_DE_DATOS=1 453 ==> GEOMETRIA_ANALITICA=1 451	conf:(1)	lift:(1.05)

Figura 98. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen algunas de las premisas obtenidas:

- 1) A nivel general, si los estudiantes aprueban Algebra Lineal, Programación I, Probabilidad e Inferencia Estadística y estructura de Datos, es muy probable que en base a estas apruebe la asignatura de Geometría Analítica, la misma que va de la mano con las otras asignaturas puesto que hacen mucho uso de la lógica y raciocinio de las anteriores.



- 2) En la regla 10, se puede visualizar que si un estudiante no trabaja y aprueba Álgebra Lineal es muy probable que apruebe Geometría analítica. Esta característica está presente en las últimas premisas, puesto que para la aprobación de dichas materias el estudiante no debe ejercer ninguna actividad laboral.

Por otra parte, se han determinado Reglas definiendo la métrica lift como predeterminado con un valor de 1.1 (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CON	LIFT			
1. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 471 ==> CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1	(0.94)	lift:(1.16)>lev:(0.1)	[61]	conv:(3.15)	
2. CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1 486 ==> ALGEBRA_LINEAL=1 PROGRAMACION_I=1	(0.91)	lift:(1.16)>lev:(0.1)	[61]	conv:(2.4)	
3. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 465 ==> CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1	(0.95)	lift:(1.15)>lev:(0.1)	[59]	conv:(3.64)	
4. CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1 495 ==> ALGEBRA_LINEAL=1 PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1	(0.9)	lift:(1.15)>lev:(0.1)	[59]	conv:(2.12)	
5. ALGEBRA_LINEAL=1 PROGRAMACION_I=1 471 ==> CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1 449	conf (0.95)	lift:(1.15)>lev:(0.1)	[59]	conv:(3.53)	
6. CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 GEOMETRIA_ANALITICA=1 495 ==> ALGEBRA_LINEAL=1 PROGRAMACION_I=1 449	conf (0.91)	lift:(1.15)>lev:(0.1)	[59]	conv:(2.24)	
7. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1 467 ==> PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 GEOMETRIA_ANALITICA=1	(0.95)	lift:(1.15)>lev:(0.1)	[58]	conv:(3.38)	
8. PROGRAMACION_I=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 GEOMETRIA_ANALITICA=1 494 ==> ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 ESTRUCTURA_DE_DATOS=1	(0.9)	lift:(1.15)>lev:(0.1)	[58]	conv:(2.12)	
9. ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1 461 ==> PROGRAMACION_I=1 GEOMETRIA_ANALITICA=1	(0.96)	lift:(1.15)>lev:(0.09)	[56]	conv:(4.07)	
10. PROGRAMACION_I=1 GEOMETRIA_ANALITICA=1 503 ==> ALGEBRA_LINEAL=1 CALCULO_INTEGRAL=1 PROBABILIDAD_E_INFERENCIA_ESTADISTICA=1 ESTRUCTURA_DE_DATOS=1	(0.88)	lift:(1.15)>lev:(0.09)	[56]	conv:(1.92)	

Figura 99. Reglas obtenidas en Weka con Lift mayor a 1

Cabe mencionar que aquí la relación entre asignaturas es alta, debido a que debe aprobar algunas para la aprobación de otras:

- 1) Si un estudiante aprueba a la asignatura de Álgebra Lineal y Programación I es muy probable que también apruebe Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de datos y Geometría Analítica
- 2) Si un estudiante aprueba Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de datos y Geometría Analítica, es muy probable que también apruebe las materias de Álgebra Lineal y Programación I.
- 3) Estudiante que apruebe Álgebra Lineal y Probabilidad e Inferencia Estadística Programación I es muy probable que también apruebe Cálculo Integral, Estructura de datos y Geometría Analítica.
- 4) Estudiante que apruebe Cálculo Integral, Estructura de datos y Geometría Analítica, es muy probable que también apruebe Álgebra Lineal, Programación I y Probabilidad e Inferencia Estadística.
- 5) Estudiante que apruebe Álgebra Lineal y Programación I, es muy probable que también apruebe Cálculo Integral, Estructura de datos y Geometría Analítica.



- 6) Si un estudiante aprueba las asignaturas de Cálculo Integral, Estructura de datos y Geometría Analítica, es muy probable que adicionalmente apruebe Algebra Lineal y Programación I.
- 7) Estudiante que apruebe Algebra Lineal, Cálculo Integral y Estructura de Datos, es muy probable que también apruebe Programación I, Probabilidad e Inferencia Estadística y Geometría Analítica.
- 8) Si un estudiante aprueba las materias de Programación I, Probabilidad e Inferencia Estadística y Geometría Analítica, es muy probable que también apruebe las asignaturas de Algebra Lineal, Cálculo Integral y Estructura de Datos.
- 9) Si un estudiante aprueba Algebra Lineal, Cálculo Integral, Probabilidad e Inferencia Estadística y Estructura de Datos, es muy probable que adicionalmente apruebe Programación I, Geometría Analítica.
- 10) Si un estudiante aprueba a Programación I, Geometría Analítica, es muy probable que también apruebe las materias de Algebra Lineal, Cálculo Integral, Probabilidad e Inferencia Estadística y Estructura de Datos.

Para este ciclo, al tomar a LIFT como métrica principal se logra identificar la dependencia que tienen unas materias de otras, debido a que en muchas de ellas se va trabajando de manera paralela desarrollándose así de manera equitativa para las mismas. Además, va muy de la mano con la confiabilidad de las premisas, ya que se cuenta con porcentajes mayores al 88%. Por otra parte, la relación entre las premisas se da de manera general entre variables académicas, y muy poco con las sociables o demográficas, al considerar una confianza del 80 % y un lift de 1.12, si se identifica la presencia de estas variables, esto también en relación con las asignaturas.

Al considerar a la CONF como métrica, se identifica lo antes mencionado, ya se muestra la presencia de los estudiantes no cuenten con una actividad laboral y así mismo la dependencia con las asignaturas.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas son muy similares a los resultados obtenidos en la herramienta weka. La relación entre las asignaturas se sigue manteniendo, sin embargo, en este caso se muestra una variable más, que es el tipo de colegio

del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que de estas bases parte a nuevos conocimientos. Esto se ve reflejado en la Figura 5.

### AssociationRules

#### Association Rules


```
[GEOMETRIA ANALITICA, tipo_colegio] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, CALCULO INTEGRAL, ALGEBRA LINEAL] (confidence: 0.798)
[tipo_colegio] --> [GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, PROGRAMACION I] (confidence: 0.799)
[GEOMETRIA ANALITICA, tipo_colegio] --> [ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I] (confidence: 0.805)
[ESTRUCTURA DE DATOS, tipo_colegio] --> [GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA
[tipo_colegio] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, PROGRAMACION I] (confidence: 0.806)
[tipo_colegio] --> [GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, CALCULO INTEGRAL] (confidence: 0.806)
[GEOMETRIA ANALITICA] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, ESTRUCTURA DE DATOS, CALCULO INTEGRAL, ALGEBRA LINEAL] (confidence: 0.807)
[GEOMETRIA ANALITICA] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, ESTRUCTURA DE DATOS, PROGRAMACION I, ALGEBRA LINEAL] (confidence: 0.808)
[PROBABILIDAD E INFERENCIA ESTADISTICA, tipo_colegio] --> [GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA
[ESTRUCTURA DE DATOS, tipo_colegio] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL] (confidence:
[tipo_colegio] --> [GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, ALGEBRA LINEAL] (confidence: 0.810)
[tipo_colegio] --> [GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, PROGRAMACION I] (confidence: 0.810)
[tipo_colegio] --> [GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, ALGEBRA LINEAL] (confidence: 0.810)
[GEOMETRIA ANALITICA, tipo_colegio] --> [CALCULO INTEGRAL, PROGRAMACION I] (confidence: 0.813)
[GEOMETRIA ANALITICA, tipo_colegio] --> [CALCULO INTEGRAL, ALGEBRA LINEAL] (confidence: 0.813)
[PROBABILIDAD E INFERENCIA ESTADISTICA, tipo_colegio] --> [GEOMETRIA ANALITICA, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL] (confidence:
[PROBABILIDAD E INFERENCIA ESTADISTICA, tipo_colegio] --> [ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL] (confidence:
[tipo_colegio] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, ALGEBRA LINEAL] (confidence: 0.814)
[tipo_colegio] --> [ESTRUCTURA DE DATOS, ALGEBRA LINEAL] (confidence: 0.814)
[tipo_colegio] --> [GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, CALCULO INTEGRAL] (confidence: 0.814)
[genero] --> [GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL] (confidence: 0.815)
[genero] --> [GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL
[GEOMETRIA ANALITICA] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, PROGRAMACION I, ALGEBRA LINEAL] (confidence: 0.815)
[GEOMETRIA ANALITICA, PROBABILIDAD E INFERENCIA ESTADISTICA, tipo_colegio] --> [ESTRUCTURA DE DATOS, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA
[GEOMETRIA ANALITICA, ESTRUCTURA DE DATOS, tipo_colegio] --> [PROBABILIDAD E INFERENCIA ESTADISTICA, CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA
[PROBABILIDAD E INFERENCIA ESTADISTICA, tipo_colegio] --> [CALCULO INTEGRAL, PROGRAMACION I, ALGEBRA LINEAL] (confidence: 0.817)
[tipo_colegio] --> [ESTRUCTURA DE DATOS, CALCULO INTEGRAL] (confidence: 0.817)
```

Figura 100. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que sean aprobadas las materias el estudiante dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, si el estado civil del alumno es soltero también aprueban todas las materias relacionadas con las matemáticas, si este pertenece al sector Urbano y la aprobación de algunas materias relacionadas de forma paralela, se muestra relación entre las asignaturas de Cálculo Diferencial y Programación I; como conclusión ante las relación que guardan estas asignaturas también se enlazan las variables de pertenecer al sector urbano, ser soltero y en este caso ante los valores asignados a la

confianza y lift no se muestra reglas con relación al tipo colegio y a la materia Estructura de Datos.



```

RelationRecord(items=frozenset({'ED_A', 'P_A', 'Urbano', 'CI_A'}), support=0.7575250836120402, ordered_statistics=[
=====
RelationRecord(items=frozenset({'PIE_A', 'P_A', 'NTrabaja', 'CI_A'}), support=0.7474916387959866, ordered_statistics=
=====
RelationRecord(items=frozenset({'PIE_A', 'NTrabaja', 'CI_A', 'Urbano'}), support=0.7324414715719063, ordered_statist
=====
RelationRecord(items=frozenset({'P_A', 'NTrabaja', 'CI_A', 'Urbano'}), support=0.7040133779264214, ordered_statistic
=====
RelationRecord(items=frozenset({'PIE_A', 'P_A', 'Urbano', 'CI_A'}), support=0.7525083612040134, ordered_statistics=[
=====
RelationRecord(items=frozenset({'ED_A', 'P_A', 'PIE_A', 'NTrabaja'}), support=0.7608695652173914, ordered_statistics=
=====
RelationRecord(items=frozenset({'ED_A', 'PIE_A', 'NTrabaja', 'Soltero'}), support=0.7006688963210702, ordered_statist
=====
RelationRecord(items=frozenset({'ED_A', 'PIE_A', 'NTrabaja', 'Urbano'}), support=0.7491638795986622, ordered_statist
=====
RelationRecord(items=frozenset({'ED_A', 'P_A', 'NTrabaja', 'Urbano'}), support=0.7190635451505016, ordered_statistic
=====
RelationRecord(items=frozenset({'ED_A', 'P_A', 'PIE_A', 'Soltero'}), support=0.7140468227424749, ordered_statistics=

```

Figura 101. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances	
0	165 ( 28%)
1	51 ( 9%)
2	207 ( 35%)
3	166 ( 28%)
4	9 ( 2%)

Figura 102. Clúster con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 2 y 3. La Figura 7, indica las características que componen a cada clúster, consecutivamente se describe cada clúster:

Final cluster centroids:

Attribute	Cluster#					
	Full Data (598.0)	0 (165.0)	1 (51.0)	2 (207.0)	3 (166.0)	4 (9.0)
edad	25	23	25	27	25	26
genero	0	1	0	0	0	0
sector_procedencia	0	0	0	0	0	1
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	0	0	1
tipo_colegio	1	1	2	2	1	2
ALGEBRA_LINEAL	1	1	0	1	1	1
PROGRAMACION_I	1	1	0	1	1	1
CALCULO_INTEGRAL	1	1	0	1	1	1
PROBABILIDAD_E_INFERENCIA_ESTADISTICA	1	1	0	1	1	1
ESTRUCTURA_DE_DATOS	1	1	0	1	1	1
GEOMETRIA_ANALITICA	1	1	1	1	1	1

Figura 103. Clústeres obtenidos por Weka.

**Clúster\_0 (28%):** Estudiantes que tengan una edad de 23 años, pertenecientes al género femenino, de estado civil soltero, provenientes de la provincia de Loja y sean del sector Urbano, que no cuenten con una actividad laboral, dedicándose únicamente a sus actividades académicas y que hayan finalizado su formación secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, es decir, Álgebra, Lineal, Programación I, Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de Datos y Geometría Analítica.

**Clúster\_1 (9%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no se encuentren realizando ninguna actividad laboral y que hayan finalizado su secundaria en una institución Fiscomisional, aprobará únicamente a Geometría Analítica, las otras materias relacionadas a las matemáticas correspondientes al segundo ciclo serán reprobadas.

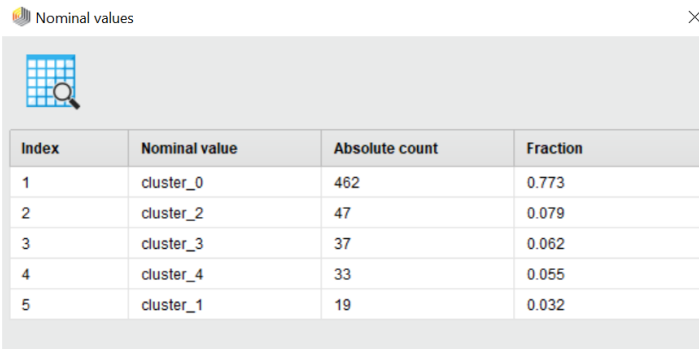
**Clúster\_2 (207%):** Estudiantes que tengan una edad de 27 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, es decir, Álgebra, Lineal, Programación I, Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de Datos y Geometría Analítica.

**Clúster\_3 (166%):** Estudiantes que tengan 25 años de edad, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan culminado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, es decir, Álgebra, Lineal, Programación I, Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de Datos y Geometría Analítica.

**Clúster\_4 (9%):** Estudiantes con una edad de 26 años, de género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Rural, que se encuentren realizando cualquier actividad laboral junto a sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, es decir, Álgebra, Lineal, Programación I, Cálculo Integral, Probabilidad e Inferencia Estadística, Estructura de Datos y Geometría Analítica.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústeres.

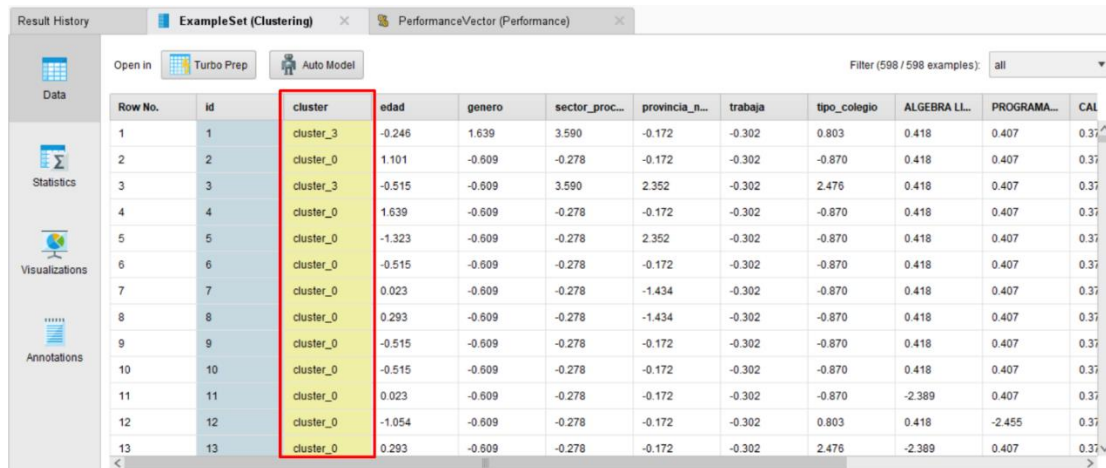


Index	Nominal value	Absolute count	Fraction
1	cluster_0	462	0.773
2	cluster_2	47	0.079
3	cluster_3	37	0.062
4	cluster_4	33	0.055
5	cluster_1	19	0.032

Figura 104. Clúster con total de instancias



Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 598 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ALGEBRA LL...	PROGRAMA...	CAL
1	1	cluster_3	-0.246	1.639	3.590	-0.172	-0.302	0.803	0.418	0.407	0.37
2	2	cluster_0	1.101	-0.609	-0.278	-0.172	-0.302	-0.870	0.418	0.407	0.37
3	3	cluster_3	-0.515	-0.609	3.590	2.352	-0.302	2.476	0.418	0.407	0.37
4	4	cluster_0	1.639	-0.609	-0.278	-0.172	-0.302	-0.870	0.418	0.407	0.37
5	5	cluster_0	-1.323	-0.609	-0.278	2.352	-0.302	-0.870	0.418	0.407	0.37
6	6	cluster_0	-0.515	-0.609	-0.278	-0.172	-0.302	-0.870	0.418	0.407	0.37
7	7	cluster_0	0.023	-0.609	-0.278	-1.434	-0.302	-0.870	0.418	0.407	0.37
8	8	cluster_0	0.293	-0.609	-0.278	-1.434	-0.302	-0.870	0.418	0.407	0.37
9	9	cluster_0	-0.515	-0.609	-0.278	-0.172	-0.302	-0.870	0.418	0.407	0.37
10	10	cluster_0	-0.515	-0.609	-0.278	-0.172	-0.302	-0.870	0.418	0.407	0.37
11	11	cluster_0	0.023	-0.609	-0.278	-0.172	-0.302	-0.870	-2.389	0.407	0.37
12	12	cluster_0	-1.054	-0.609	-0.278	-0.172	-0.302	0.803	0.418	-2.455	0.37
13	13	cluster_0	0.293	-0.609	-0.278	-0.172	-0.302	2.476	-2.389	0.407	0.37

Figura 105. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo en este caso el clúster 0 con más instancias agrupadas.

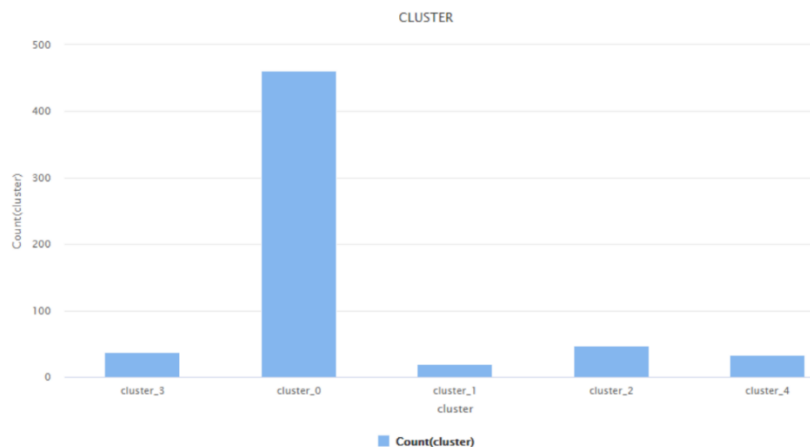


Figura 106. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, al ejecutarlo se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

### k-Means - Centroid Table

Cluster	ALGEBR...	CALCUL...	edad	ESTRUC...	genero	GEOME...	PROBA...	PROGR...	provinci...	sector_...	tipo_col...	trabaja
Cluster 0	1.000	0.973	27.169	0.993	0.306	1	0.991	0.966	11.467	0.074	1.503	-0
Cluster 1	0.917	0.833	26.938	0.937	0.188	0.979	0.958	0.896	11.583	0.063	1.438	1.000
Cluster 2	0.053	0.053	25.789	0.158	0.158	0	0.105	0.053	11.211	0.158	1.632	0.053
Cluster 3	0.400	0.267	24.422	0.289	0.178	1	0.333	0.044	11.844	0.067	1.711	0.022
Cluster 4	0.024	0.951	27.366	0.927	0.146	0.780	0.951	0.902	12.171	0.024	1.537	0

Figura 107. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas tales como Álgebra Lineal, Cálculo Integral, Estructura de Datos, Geometría Analítica, Probabilidad e Inferencia Estadística y Programación I, todas pertenecientes al segundo ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas tales como Álgebra Lineal, Cálculo Integral, Estructura de Datos, Geometría Analítica, Probabilidad e Inferencia Estadística y Programación I, todas pertenecientes al segundo ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y realiza una actividad laboral extra a sus actividades académicas.

**Clúster 2:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas, estas son Álgebra Lineal, Cálculo Integral, Estructura de Datos, Geometría Analítica, Probabilidad e Inferencia Estadística y Programación I, si tiene una edad de 25 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 3:** Estudiante aprueba únicamente la asignatura de Geometría Analítica y reprueba a la mayoría de las asignaturas relacionadas a las matemáticas tales como Álgebra Lineal, Cálculo Integral, Estructura de Datos, Probabilidad e Inferencia Estadística y Programación I, todas pertenecientes al segundo ciclo, si tiene una edad de 24 años, es de género masculino, oriundo de la provincia



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Cálculo Integral, Estructura de Datos, Geometría Analítica, Probabilidad e Inferencia Estadística y Programación I, excepto Algebra Lineal, todas pertenecientes al primer ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Tipo de colegio** y **Género**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.



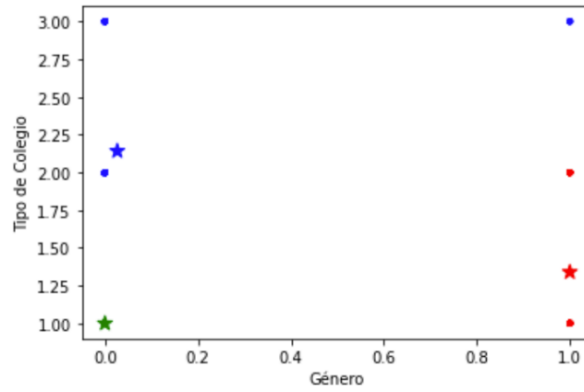


Figura 108. Relación Género - Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Programación I**, donde aprueban a la asignatura estudiantes con edades de 22, 27 y 32 años.

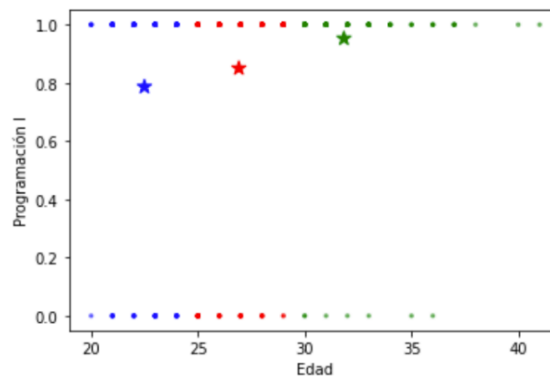


Figura 109. Relación Edad – Programación I

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Programación I**, donde se observa que el género masculino aprueba y reprueba Programación I, mientras que hay un solo clúster que engloba al género femenino para la aprobación de dicha asignatura.

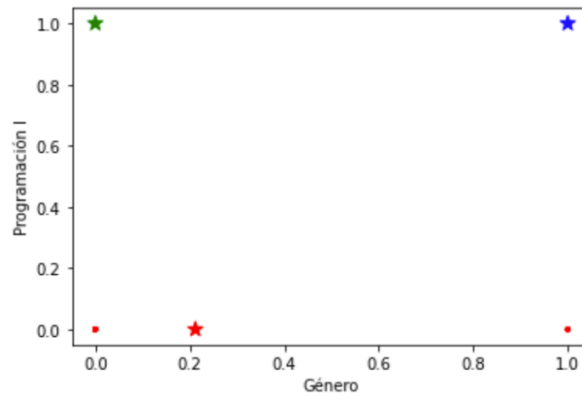


Figura 110. Relación Género - Programación I

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Cálculo Diferencial** y **Programación I**, donde se puede ver que ambas se desarrollan de manera paralela, en algunos casos unos dependen de la otra, es decir, si aprueba Cálculo diferencial reprueba Programación y viceversa; si se trabajan juntas o bien son aprobadas o reprobadas.

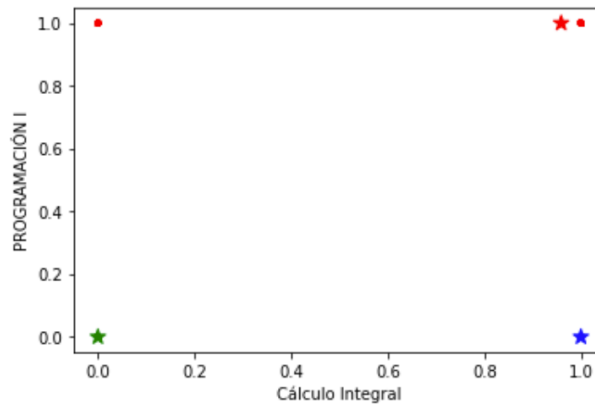


Figura 111. Relación Cálculo Diferencial – Programación I

## Anexo 6. Ingeniería en Sistemas – Tercer ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **tercer ciclo de la carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

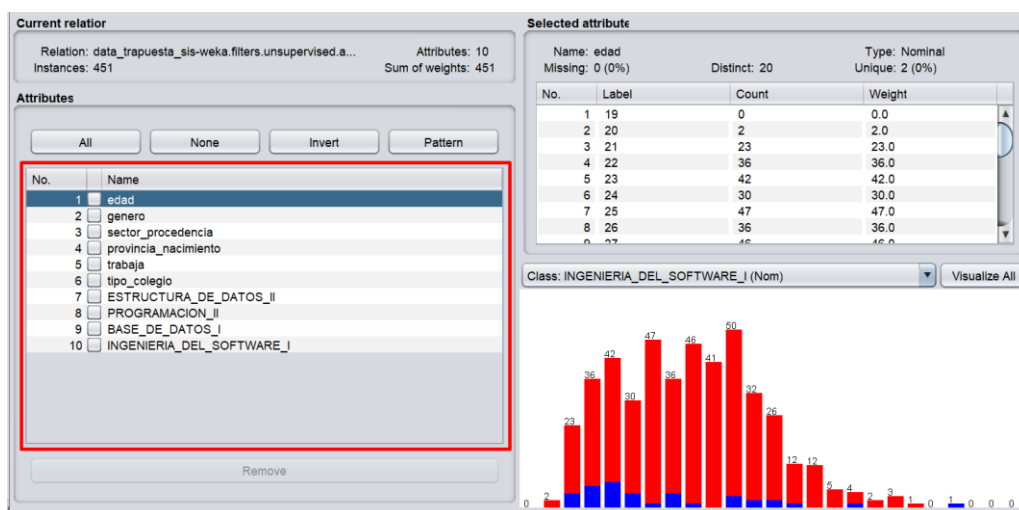


Figura 112. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 384 ==> PROGRAMACION_II=1 INGENIERIA_DEL_SO
177. trabaja=0 405 ==> sector_procedencia=0 371 <conf:(0.92)> lift:(1) lev:(-0) [0] conv:(0.95)
178. trabaja=0 405 ==> ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 371 <conf:(0.92)> lift:(1) lev:(-0) [-1] co
179. PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 393 ==> sector_procedencia=0 BASE_DE_DATOS_I=1 360 <conf:(
180. trabaja=0 ESTRUCTURA_DE_DATOS_II=1 376 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1
181. ESTRUCTURA_DE_DATOS_II=1 420 ==> sector_procedencia=0 BASE_DE_DATOS_I=1 384 <conf:(0.91)> lift:(1.04) le
182. trabaja=0 405 ==> PROGRAMACION_II=1 370 <conf:(0.91)> lift:(1.01) lev:(0.01) [2] conv:(1.05)
183. trabaja=0 405 ==> INGENIERIA_DEL_SOFTWARE_I=1 370 <conf:(0.91)> lift:(1) lev:(-0) [-1] conv:(0.92)
184. BASE_DE_DATOS_I=1 427 ==> PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 390 <conf:(0.91)> lift:(1.05) le
185. sector_procedencia=0 414 ==> PROGRAMACION_II=1 378 <conf:(0.91)> lift:(1.01) lev:(0.01) [2] conv:(1.04)
186. INGENIERIA_DEL_SOFTWARE_I=1 414 ==> sector_procedencia=0 BASE_DE_DATOS_I=1 378 <conf:(0.91)> lift:(1.04)
187. sector_procedencia=0 414 ==> BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 378 <conf:(0.91)> lift:(1.01)
188. trabaja=0 ESTRUCTURA_DE_DATOS_II=1 376 ==> sector_procedencia=0 BASE_DE_DATOS_I=1 343 <conf:(0.91)> lift
189. ESTRUCTURA_DE_DATOS_II=1 420 ==> PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 383 <conf:(0.91)> lift:(1
190. sector_procedencia=0 BASE_DE_DATOS_I=1 395 ==> PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 360 <conf:(
191. sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 388 ==> PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 353
192. PROGRAMACION_II=1 409 ==> sector_procedencia=0 BASE_DE_DATOS_I=1 372 <conf:(0.91)> lift:(1.04) lev:(0.03
193. sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 388 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SO
194. ESTRUCTURA_DE_DATOS_II=1 420 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 381 <co
195. ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 396 ==> trabaja=0 359 <conf:(0.91)> lift:(1.01) lev:(0.01) [3
196. ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 393 ==> trabaja=0 356 <conf:(0.91)> lift:(1
197. PROGRAMACION_II=1 BASE_DE_DATOS_I=1 402 ==> sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 364 <conf:(0.9
198. PROGRAMACION_II=1 409 ==> trabaja=0 370 <conf:(0.9)> lift:(1.01) lev:(0.01) [2] conv:(1.04)
199. ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 383 ==> trabaja=0 346 <conf:(0.9
200. PROGRAMACION_II=1 BASE_DE_DATOS_I=1 402 ==> trabaja=0 363 <conf:(0.9)> lift:(1.01) lev:(0) [2] conv:(1.0
  
```

Figura 113. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza mayor a 0.9 y lift mayor a 1 respectivamente, se obtiene un total de 16 reglas (Ver Figura 3).

	A	B	C	D	E	F
	REGLAS	CONF	LIFT			
37	sector_procedencia=0 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 362 ==> ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 352	conf:(0.97)>	lift:(1.06)	lev:(0.04)	[18]	conv:(2.63)
40	trabaja=0 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 354 ==> ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 344	conf:(0.97)>	lift:(1.06)	lev:(0.04)	[18]	conv:(2.57)
43	PROGRAMACION_II=1 BASE_DE_DATOS_I=1 402 ==> INGENIERIA_DEL_SOFTWARE_I=1 390	conf:(0.97)>	lift:(1.06)	lev:(0.05)	[20]	conv:(2.54)
44	ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 393 ==> INGENIERIA_DEL_SOFTWARE_I=1 381	conf:(0.97)>	lift:(1.06)	lev:(0.04)	[20]	conv:(2.48)
46	trabaja=0 ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 355 ==> PROGRAMACION_II=1 344	conf:(0.97)>	lift:(1.07)	lev:(0.05)	[22]	conv:(2.75)
57	trabaja=0 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 364 ==> PROGRAMACION_II=1 351	conf:(0.96)>	lift:(1.06)	lev:(0.05)	[20]	conv:(2.42)
58	trabaja=0 ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 359 ==> PROGRAMACION_II=1 346	conf:(0.96)>	lift:(1.06)	lev:(0.05)	[20]	conv:(2.39)
62	ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 396 ==> BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 381	conf:(0.96)>	lift:(1.06)	lev:(0.05)	[22]	conv:(2.36)
63	sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 366 ==> BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 352	conf:(0.96)>	lift:(1.06)	lev:(0.05)	[20]	conv:(2.33)
67	ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 397 ==> PROGRAMACION_II=1 381	conf:(0.96)>	lift:(1.06)	lev:(0.05)	[20]	conv:(2.17)
68	trabaja=0 ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 371 ==> PROGRAMACION_II=1 356	conf:(0.96)>	lift:(1.06)	lev:(0.04)	[19]	conv:(2.16)
72	PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 359 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 344	conf:(0.96)>	lift:(1.08)	lev:(0.05)	[24]	conv:(2.44)
73	trabaja=0 ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 359 ==> BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 344	conf:(0.96)>	lift:(1.06)	lev:(0.04)	[19]	conv:(2.14)
77	trabaja=0 INGENIERIA_DEL_SOFTWARE_I=1 370 ==> PROGRAMACION_II=1 354	conf:(0.96)>	lift:(1.06)	lev:(0.04)	[18]	conv:(2.03)
95	ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 401 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 381	conf:(0.95)>	lift:(1.07)	lev:(0.05)	[23]	conv:(2.07)
98	sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 371 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 352	conf:(0.95)>	lift:(1.06)	lev:(0.05)	[21]	conv:(2.02)
99	trabaja=0 INGENIERIA_DEL_SOFTWARE_I=1 370 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 351	conf:(0.95)>	lift:(1.06)	lev:(0.05)	[21]	conv:(2.01)
103	PROGRAMACION_II=1 BASE_DE_DATOS_I=1 402 ==> ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 381	conf:(0.95)>	lift:(1.07)	lev:(0.05)	[23]	conv:(2.03)
104	trabaja=0 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 363 ==> ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 344	conf:(0.95)>	lift:(1.07)	lev:(0.05)	[21]	conv:(2.01)
106	trabaja=0 ESTRUCTURA_DE_DATOS_II=1 376 ==> PROGRAMACION_II=1 BASE_DE_DATOS_I=1 356	conf:(0.95)>	lift:(1.06)	lev:(0.05)	[20]	conv:(1.95)
107	sector_procedencia=0 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 372 ==> ESTRUCTURA_DE_DATOS_II=1 INGENIERIA_DEL_SOFTWARE_I=1 352	conf:(0.95)>	lift:(1.06)	lev:(0.05)	[21]	conv:(1.96)
110	trabaja=0 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 364 ==> ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 344	conf:(0.95)>	lift:(1.08)	lev:(0.05)	[24]	conv:(2.11)

Figura 114. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante proveniente del sector Urbano y aprueba Programación II e Ingeniería del software I, es muy probable que también apruebe Estructura de Datos II y Base de datos I.



- 2) Si un estudiante no ejerce ninguna actividad laboral, aprueba Programación II e Ingeniería del software I, es muy probable que también apruebe Estructura de Datos II y Base de datos I.
- 3) Estudiante que apruebe Programación II y Base de datos I muy probablemente también aprobará Ingeniería del software I.
- 4) Estudiante que apruebe Estructura de Datos II, Programación II y Base de datos I muy probablemente también aprobará Ingeniería del software I.
- 5) Estudiante que no efectúe ninguna actividad laboral y que apruebe Ingeniería del software, Estructura de Datos II y Base de datos I muy probablemente también aprobará Programación II.
- 6) Estudiante que no efectúe ninguna actividad laboral y que apruebe Ingeniería del software y Base de datos, I muy probablemente también aprobará Programación II.
- 7) Estudiante que no efectúe ninguna actividad laboral y que apruebe Ingeniería del software y Estructura de Datos II, muy probablemente también aprobará Programación II.
- 8) Estudiante que apruebe Estructura de Datos II y Base de datos I muy probablemente también aprobará Programación II e Ingeniería del software.
- 9) Estudiante que procedente del sector Urbano y que apruebe Ingeniería del software, Estructura de Datos II y Programación II, probablemente también aprobará Base de datos I e Ingeniería del software I.
- 10) Estudiante que apruebe Estructura de Datos II, Base de datos I e Ingeniería del software muy probablemente también aprobará Programación II.
- 11) Estudiante que no ejerza ninguna actividad laboral, apruebe Estructura de Datos II y Base de dato I, probablemente apruebe Programación II.
- 12) Estudiante que no ejerza ninguna actividad laboral, apruebe Estructura de Datos II e Ingeniería del software I probablemente apruebe Programación II y Base de Datos I.
- 13) Estudiante que no ejerza ninguna actividad laboral, apruebe Estructura de Datos II y Programación II probablemente apruebe Base de Datos I e Ingeniería del Software I.

- 14) Estudiante que no ejerza ninguna actividad laboral y apruebe Ingeniería del software aprobará también Programación II.
- 15) Estudiante que apruebe Estructura de datos e Ingeniería del software II aprobará Programación II y Base de Datos I.
- 16) Estudiante que pertenezca al sector Urbano y apruebe Estructura de datos II e Ingeniería del software I aprobará también Programación II y Base de datos I.
- 17) Estudiante que no ejerza ninguna actividad laboral y apruebe Ingeniería del software I muy probablemente se aprobará Programación II y Base de datos I.
- 18) Estudiante que no ejerza ninguna actividad laboral y apruebe Estructura de Datos II, muy probablemente se aprobará Programación II y Base de datos I.
- 19) Estudiante que no ejerza ninguna actividad laboral, y apruebe Base de Datos e Ingeniería del software I, posiblemente aprobará Estructura de Datos II y Programación II.

Por otra parte, se han determinado Reglas definiendo Lift como predeterminado con valor de 1.1 (Ver Figura 4) Considerando que en este caso la confianza abarca únicamente un 75% en comparación a otros ciclos que trabajan con una confianza mayor al 90%. A continuación, se tienen los siguientes resultados:

REGLAS	CONF	LIFT		
95. tipo_colegio=2 ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 163 ==> genero=0 sector_procedencia=0 BASE_DE_DATOS_I=1 124 conf:	0.76)	lift:(1.21)>	lev:(0.05)	[21]
100. tipo_colegio=2 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 166 ==> genero=0 sector_procedencia=0 ESTRUCTURA_DE_DATOS_II=1 124 conf:	0.75)	lift:(1.21)>	lev:(0.05)	[21]
156. tipo_colegio=2 ESTRUCTURA_DE_DATOS_II=1 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 155 ==> genero=0 sector_procedencia=0 BASE_DE_DATOS_I=1 117 conf:	0.75)	lift:(1.2)>	lev:(0.04)	[19]
160. trabaja=0 tipo_colegio=2 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 143 ==> sector_procedencia=0 provincia_nacimiento=11 ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 114 conf:	0.8)	lift:(1.2)>	lev:(0.04)	[18]
186. trabaja=0 tipo_colegio=2 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 143 ==> sector_procedencia=0 provincia_nacimiento=11 ESTRUCTURA_DE_DATOS_II=1 115 conf:	0.8)	lift:(1.19)>	lev:(0.04)	[18]
195. tipo_colegio=2 PROGRAMACION_II=1 INGENIERIA_DEL_SOFTWARE_I=1 159 ==> sector_procedencia=0 provincia_nacimiento=11 ESTRUCTURA_DE_DATOS_II=1 BASE_DE_DATOS_I=1 126 conf:	0.79)	lift:(1.19)>	lev:(0.04)	[20]
199. trabaja=0 tipo_colegio=2 PROGRAMACION_II=1 BASE_DE_DATOS_I=1 INGENIERIA_DEL_SOFTWARE_I=1 142 ==> sector_procedencia=0 provincia_nacimiento=11 ESTRUCTURA_DE_DATOS_II=1 114 conf:	0.8)	lift:(1.19)>	lev:(0.04)	[18]
200. tipo_colegio=2 PROGRAMACION_II=1 168 ==> genero=0 sector_procedencia=0 BASE_DE_DATOS_I=1 126 conf:	0.75)	lift:(1.19)>	lev:(0.04)	[20]

Figura 115. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Estructura de datos II y Programación II, es muy probable que sea de sexo masculino, proveniente del sector Urbano y que también apruebe Base de datos I.
- 2) Si un estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Base de datos I y Programación II, es muy probable que sea de sexo masculino, proveniente del sector Urbano y que también apruebe Estructura de datos II.





- 3) Si un estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Estructura de datos II y Programación II, es muy probable que sea de sexo masculino, proveniente del sector Urbano y que también apruebe Base de datos I.
- 4) Si un estudiante no tiene ninguna actividad laboral, ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Programación II e Ingeniería del software II, es muy probable que provenga del sector Urbano, pertenezca a la provincia de Loja que también apruebe Estructura de datos II y Base de datos I.
- 5) Si un estudiante no tiene ninguna actividad laboral, ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Programación II e Ingeniería del software II, es muy probable que provenga del sector Urbano, pertenezca a la provincia de Loja que también apruebe Estructura de datos II.
- 6) Si estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Programación II e Ingeniería del software II, es muy probable que provenga del sector Urbano, pertenezca a la provincia de Loja y también que apruebe la asignatura de Estructura de datos II y también Base de datos I.
- 7) Si estudiante cumple con una actividad laboral, culmina sus estudios secundarios en un colegio Fiscomisional y aprueba Programación II, Base de datos e Ingeniería del software II, es muy probable que provenga del sector Urbano, pertenezca a la provincia de Loja y también que apruebe la asignatura de Estructura de datos II.
- 8) Estudiante que culmina sus estudios secundarios en un colegio Fiscomisional y aprueba Programación II es muy probable que sea de sexo masculino, que sea procedente del sector Urbano y que apruebe Base de datos I.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro

de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es muy baja, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo.

## Resultados obtenidos por Algoritmo A priori – RapidMiner

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, sin embargo, en este caso se muestra una variable más, que es el tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 21 a 25 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

Association Rules

```
[tipo_colegio] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.80)
[tipo_colegio] --> [BASE DE DATOS I, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.825)
[genero] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.827)
[genero] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.827)
[tipo_colegio] --> [INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.830)
[genero] --> [ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.835)
[genero] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.835)
[tipo_colegio] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.835)
[tipo_colegio] --> [ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.84)
[tipo_colegio, edad] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.842)
[tipo_colegio, edad] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.842)
[tipo_colegio] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I] (confidence: 0.844)
[BASE DE DATOS I, tipo_colegio] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.854)
[tipo_colegio] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I] (confidence: 0.854)
[tipo_colegio, edad] --> [ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.855)
[tipo_colegio, edad] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, PROGRAMACION II] (confidence: 0.855)
[genero] --> [BASE DE DATOS I, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.858)
[tipo_colegio] --> [BASE DE DATOS I, PROGRAMACION II] (confidence: 0.863)
[genero] --> [BASE DE DATOS I, PROGRAMACION II] (confidence: 0.866)
[genero] --> [INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.87)
[tipo_colegio, edad] --> [ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I] (confidence: 0.868)
[tipo_colegio, edad] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I] (confidence: 0.868)
[genero] --> [PROGRAMACION II] (confidence: 0.874)
[genero] --> [BASE DE DATOS I, ESTRUCTURA DE DATOS II, INGENIERIA DEL SOFTWARE I] (confidence: 0.874)
[tipo_colegio] --> [PROGRAMACION II] (confidence: 0.877)
[tipo_colegio] --> [BASE DE DATOS I, INGENIERIA DEL SOFTWARE I] (confidence: 0.877)
[ESTRUCTURA DE DATOS II, tipo_colegio] --> [BASE DE DATOS I, INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.880)
[tipo_colegio, edad] --> [INGENIERIA DEL SOFTWARE I, PROGRAMACION II] (confidence: 0.882)
```

Figura 116. Reglas de asociación en RapidMiner

Muchas de las relaciones que existe en las premisas es la presencia del antecedente a tipo de colegio y la edad, y como consecuente la aprobación de





las asignaturas, esto presenta mucha lógica debido a que las bases impartidas en instituciones Fiscales y Fiscomisionales (de ser el caso y dependiendo de la ubicación de la misma) no son las suficientes y necesarias para la continuidad en la formación universitaria, sin embargo si bien es cierto que, esto presenta un obstáculo para estas materias, se debe considerar que los estudiante ya se encuentran en un tercer ciclo y ya debió obtener algunas bases en los dos primeros ciclos que permitan al estudiante forjarse para los posteriores.

### **Resultados obtenidos por Algoritmo A priori – Python**

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que sean aprobadas las materias el estudiante dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, si el estado civil del alumno es soltero también aprueban todas las materias relacionadas con las matemáticas, si este pertenece al sector Urbano y la aprobación de algunas materias relacionadas de forma paralela, se muestra relación entre las asignaturas Programación II y Estructura de Datos II, en ella junto con otras características son necesarias para la aprobación de ambas; como conclusión ante las relación que guardan estas asignaturas también se enlazan las variables de pertenecer al sector urbano, ser soltero y en este caso ante los valores asignados a la confianza y lift no se muestra reglas con relación al tipo colegio y a la materia Base de Datos I.

```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardar...
RAM Disco Editando
+ Código + Texto
RelationRecord(items=frozenset({'PII_A', 'NTrabaja', 'EDII_', 'Urbano'}), support=0.7339246119733924, ordered_statist
RelationRecord(items=frozenset({'Urbano', 'PII_A', 'EDII_', 'Soltero'}), support=0.7272727272727273, ordered_statist
RelationRecord(items=frozenset({'IS_A', 'NTrabaja', 'PII_A', 'Soltero'}), support=0.70509977827051, ordered_statist
RelationRecord(items=frozenset({'IS_A', 'NTrabaja', 'PII_A', 'Urbano'}), support=0.720620842572062, ordered_statist
RelationRecord(items=frozenset({'IS_A', 'Urbano', 'PII_A', 'Soltero'}), support=0.7184035476718403, ordered_statist
RelationRecord(items=frozenset({'IS_A', 'NTrabaja', 'PII_A', 'BD_A', 'EDII_'}), support=0.7627494456762749, ordered_
RelationRecord(items=frozenset({'IS_A', 'NTrabaja', 'BD_A', 'EDII_', 'Soltero'}), support=0.7095343680709535, orde
RelationRecord(items=frozenset({'IS_A', 'NTrabaja', 'Urbano', 'BD_A', 'EDII_'}), support=0.7272727272727273, ordere
RelationRecord(items=frozenset({'IS_A', 'PII_A', 'BD_A', 'EDII_', 'Soltero'}), support=0.7583148558758315, ordered_s
RelationRecord(items=frozenset({'IS_A', 'PII_A', 'Urbano', 'BD_A', 'EDII_'}), support=0.7804878048780488, ordered_st

```

Figura 117. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 7.

Clustered Instances	
0	17 ( 4%)
1	181 ( 40%)
2	10 ( 2%)
3	217 ( 48%)
4	26 ( 6%)

Figura 118. clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 1 y 3. La Figura 8, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

```

Final cluster centroids:
Attribute          Full Data      Cluster#
                   (451.0)        0          1          2          3          4
=====
edad               29             23         24         22         25         22
genero             0              0          0          1          0          0
sector_procedencia 0              0          0          0          0          0
provincia_nacimiento 11            11         11         11         11         11
trabaja            0              0          0          0          0          0
tipo_colegio       1              2          2          1          1          2
ESTRUCTURA_DE_DATOS_II 1            0          1          0          1          1
PROGRAMACION_II   1              0          1          0          1          1
BASE_DE_DATOS_I   1              0          1          0          1          1
INGENIERIA_DEL_SOFTWARE_I 1            0          1          0          1          1

```

Figura 119. Clúster obtenidos por Weka.



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

**Clúster\_0 (4%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, reprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I.

**Clúster\_1 (40%):** Estudiantes que tengan una edad de 24 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I.

**Clúster\_2 (2%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, reprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I.

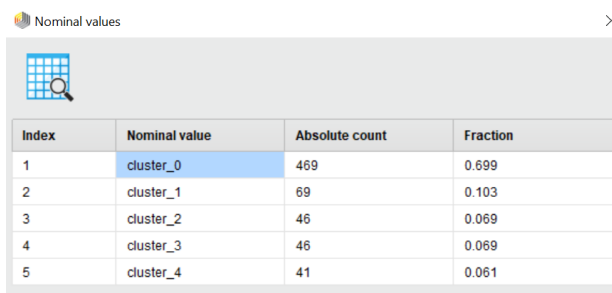
**Clúster\_3 (48%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I.

**Clúster\_4 (6%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución

Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

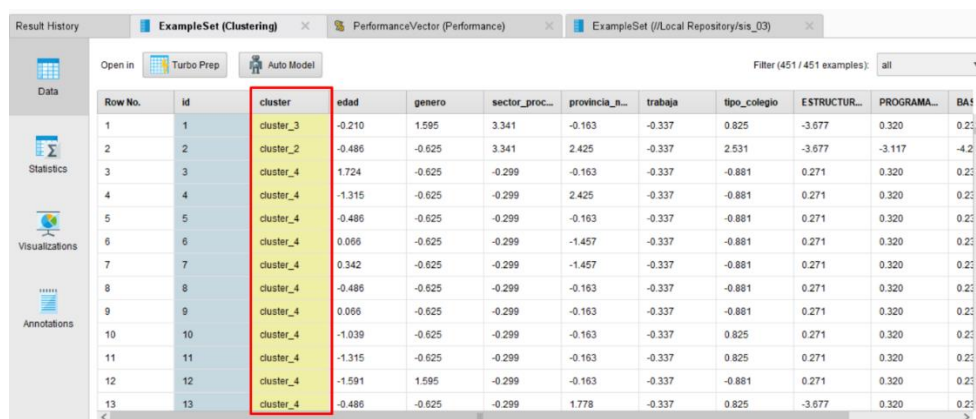
La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	469	0.699
2	cluster_1	69	0.103
3	cluster_2	46	0.069
4	cluster_3	46	0.069
5	cluster_4	41	0.061

Figura 120. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 451 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ESTRUCTUR...	PROGRAMA...	BAI
1	1	cluster_3	-0.210	1.595	3.341	-0.163	-0.337	0.825	-3.677	0.320	0.22
2	2	cluster_2	-0.486	-0.625	3.341	2.425	-0.337	2.531	-3.677	-3.117	-4.2
3	3	cluster_4	1.724	-0.625	-0.299	-0.163	-0.337	-0.881	0.271	0.320	0.22
4	4	cluster_4	-1.315	-0.625	-0.299	2.425	-0.337	-0.881	0.271	0.320	0.22
5	5	cluster_4	-0.486	-0.625	-0.299	-0.163	-0.337	-0.881	0.271	0.320	0.22
6	6	cluster_4	0.066	-0.625	-0.299	-1.457	-0.337	-0.881	0.271	0.320	0.22
7	7	cluster_4	0.342	-0.625	-0.299	-1.457	-0.337	-0.881	0.271	0.320	0.22
8	8	cluster_4	-0.486	-0.625	-0.299	-0.163	-0.337	-0.881	0.271	0.320	0.22
9	9	cluster_4	0.066	-0.625	-0.299	-0.163	-0.337	-0.881	0.271	0.320	0.22
10	10	cluster_4	-1.039	-0.625	-0.299	-0.163	-0.337	0.825	0.271	0.320	0.22
11	11	cluster_4	-1.315	-0.625	-0.299	-0.163	-0.337	0.825	0.271	0.320	0.22
12	12	cluster_4	-1.591	1.595	-0.299	-0.163	-0.337	-0.881	0.271	0.320	0.22
13	13	cluster_4	-0.486	-0.625	-0.299	1.778	-0.337	0.825	-3.677	0.320	0.22

Figura 121. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 4 con más instancias agrupadas.

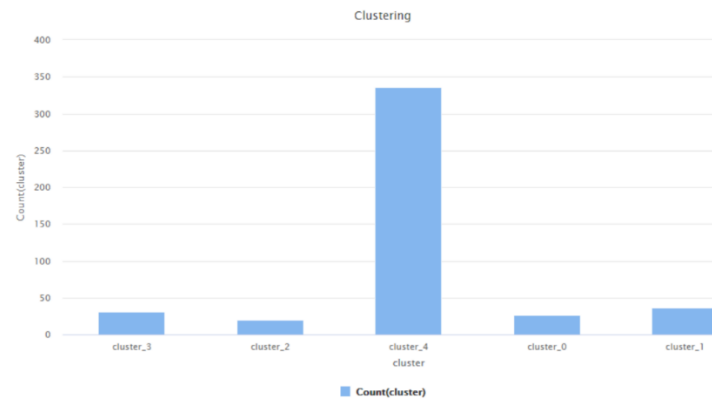


Figura 122. Representación gráfica de los clústers obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	BASE DE D...	edad	ESTRUCTU...	genero	INGENIERI...	PROGRAM...	provincia_...	sector_pr...	tipo_colegio	trabaja
Cluster 0	1	26.628	0.953	0.186	0.977	0.860	11.884	0	1.419	1
Cluster 1	0.120	25.800	0.120	0.400	0.240	0.200	11.320	0.160	1.760	0
Cluster 2	0.996	25.390	0.991	0.325	0.947	0.969	10.855	-0	1.329	-0
Cluster 3	1	29.451	0.967	0.238	0.975	0.943	12.590	0	1.861	0
Cluster 4	0.970	27.182	0.970	0.182	0.939	0.939	11.606	1.000	1.485	0.091

Figura 123. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I, todas pertenecientes al tercer ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y a parte de sus actividades académicas ejerce alguna actividad laboral.

**Clúster 1:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I, todas pertenecientes al tercer ciclo, si tiene una edad de 25 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y



únicamente se dedica a sus actividades académicas, es decir, no ejerce alguna actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I, todas pertenecientes al tercer ciclo, si tiene una edad de 25 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no ejerce alguna actividad laboral.

**Clúster 3:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I, todas pertenecientes al tercer ciclo, si tiene una edad de 29 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no ejerce alguna actividad laboral.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estructura de datos II, Programación II, Base de datos I e Ingeniería del software I, todas pertenecientes al tercer ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Rural, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no ejerce alguna actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.

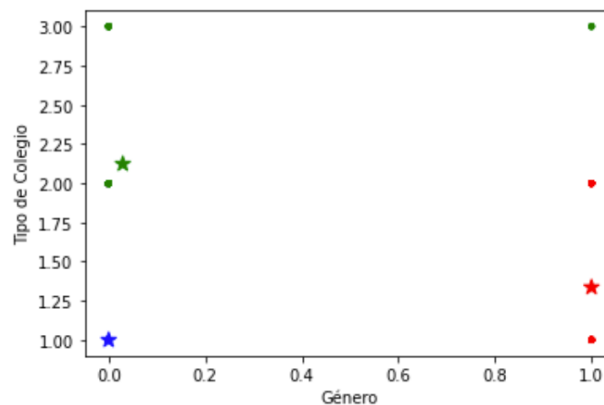


Figura 124. Relación Género - Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Programación II**, en donde los estudiantes aprueban a la asignatura si su edad es de 23, 27 y 32 años.

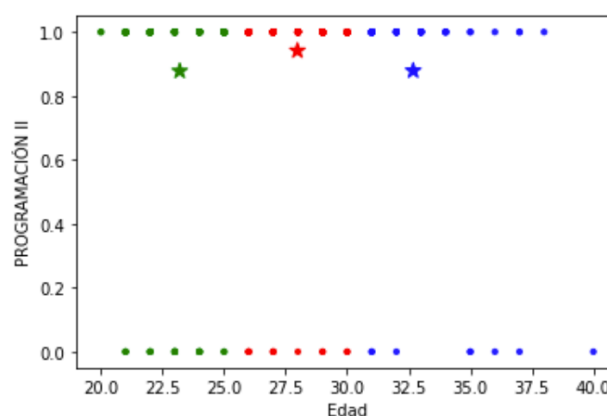


Figura 125. Relación Edad – Programación II

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Programación II**, donde se observa que un grupo mínimo de estudiantes del género femenino aprueban la materia, por otro lado, el género masculino también



aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que probabilísticamente tiendan a reprobar la materia de Programación II.

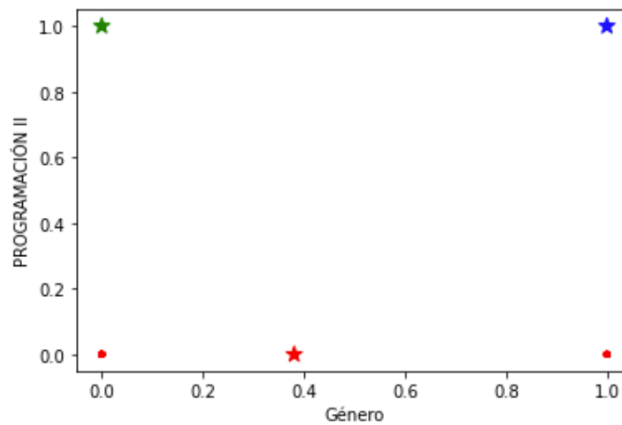
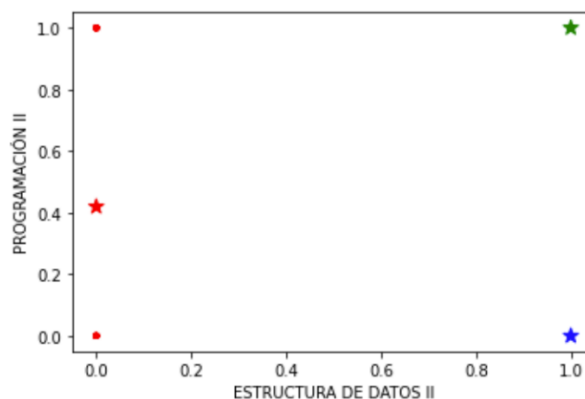


Figura 126. Relación Género – Programación II

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Estructura de Datos II** y **Programación I**, donde se mantiene el mismo comportamiento de ciclos anteriores, en este caso se da el caso que ambas asignaturas sean aprobadas, mientras que si aprueba Estructura de Datos II reprueba Programación II; por consiguiente, existe una alta probabilidad de que pierdan ambas materias considerando a estas una leve posibilidad que apruebe únicamente a Programación II.





## Anexo 7. Ingeniería en Sistemas – Cuarto ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **cuarto ciclo de la carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

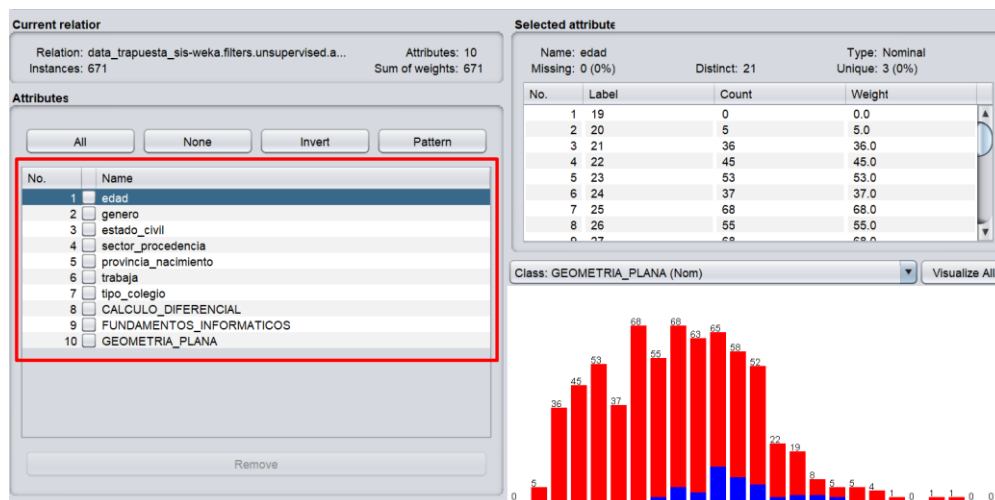


Figura 127. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. trabaja=0 CONTABILIDAD_GENERAL=1 385 ==> sector_procedencia=0 354 <conf:(0.92)> lift:(0.99) lev:(-0.01) [-1
177. sector_procedencia=0 trabaja=0 385 ==> CONTABILIDAD_GENERAL=1 354 <conf:(0.92)> lift:(0.99) lev:(-0.01)
178. METODOLOGIA_DE_LA_INVESTIGACION=1 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 372 ==> sector_pr
179. trabaja=0 METODOLOGIA_DE_LA_INVESTIGACION=1 CONTABILIDAD_GENERAL=1 384 ==> sector_procedencia=0 353 <con
180. provincia_nacimiento=11 METODOLOGIA_DE_LA_INVESTIGACION=1 361 ==> DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 CONT
181. METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 373 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE B
182. METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 373 ==> sector_procedencia=0 METODOLOGIA_DE_LA_INVE
183. trabaja=0 CONTABILIDAD_GENERAL=1 385 ==> sector_procedencia=0 METODOLOGIA_DE_LA_INVESTIGACION=1 353 <con
184. sector_procedencia=0 trabaja=0 385 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 CONTABILIDAD_GENERAL=1 353 <con
185. CONTABILIDAD_GENERAL=1 429 ==> sector_procedencia=0 METODOLOGIA_DE_LA_INVESTIGACION=1 393 <conf:(0.92)>
186. METODOLOGIA_DE_LA_INVESTIGACION=1 METODOLOGIA_DE_LA_PROGRAMACION=1 381 ==> sector_procedencia=0 DISEÑO_Y_G
187. sector_procedencia=0 427 ==> DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 CONTABILIDAD_GENERAL=1 391 <conf:(0.92)
188. provincia_nacimiento=11 365 ==> sector_procedencia=0 METODOLOGIA_DE_LA_INVESTIGACION=1 334 <conf:(0.92)>
189. trabaja=0 METODOLOGIA_DE_LA_INVESTIGACION=1 410 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE DATOS
190. provincia_nacimiento=11 METODOLOGIA_DE_LA_INVESTIGACION=1 361 ==> sector_procedencia=0 DISEÑO_Y_GESTION DE
191. METODOLOGIA_DE_LA_PROGRAMACION=1 382 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 349 <
192. METODOLOGIA_DE_LA_PROGRAMACION=1 382 ==> sector_procedencia=0 METODOLOGIA_DE_LA_INVESTIGACION=1 DISEÑO_Y_G
193. trabaja=0 416 ==> DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 CONTABILIDAD_GENERAL=1 380 <conf:(0.91)> lift:(1)
194. METODOLOGIA_DE_LA_INVESTIGACION=1 CONTABILIDAD_GENERAL=1 427 ==> sector_procedencia=0 DISEÑO_Y_GESTION DE
195. sector_procedencia=0 427 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 CONTABI
196. METODOLOGIA_DE_LA_INVESTIGACION=1 455 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 415
197. trabaja=0 CONTABILIDAD_GENERAL=1 385 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 351 <
198. sector_procedencia=0 trabaja=0 385 ==> DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 CONTABILIDAD_GENERAL=1 351 <
199. trabaja=0 METODOLOGIA_DE_LA_INVESTIGACION=1 CONTABILIDAD_GENERAL=1 384 ==> sector_procedencia=0 DISEÑO_Y_G
200. CONTABILIDAD_GENERAL=1 429 ==> sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 391 <conf:(0.91)
  
```

Figura 128. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza igual a 1 y lift mayor a 1, se obtiene un total de 12 premisas (Ver Figura 3).

REGLAS	CONF	LIFT	EV	D	E	F
1. DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 376 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 376	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.7)
2. DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 369 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 369	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.58)
3. provincia_nacimiento=11 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_INVESTIGACION=1 355	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.39)
4. sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 349 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 349	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.29)
5. sector_procedencia=0 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 342 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 342	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.18)
6. trabaja=0 METODOLOGIA_DE_LA_PROGRAMACION=1 341 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 341	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.17)
7. trabaja=0 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 337 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 337	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.11)
8. trabaja=0 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 333 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 333	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.05)
9. provincia_nacimiento=11 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 CONTABILIDAD_GENERAL=1 331 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 331	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5.02)
10. sector_procedencia=0 provincia_nacimiento=11 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 330 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 330	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5)
11. trabaja=0 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 330 ==> METODOLOGIA_DE_LA_INVESTIGACION=1 330	conf(1)	lift(1.02)	evr(0.01)	[5]		convr(5)
26. sector_procedencia=0 METODOLOGIA_DE_LA_INVESTIGACION=1 METODOLOGIA_DE_LA_PROGRAMACION=1 CONTABILIDAD_GENERAL=1 343 ==> DISEÑO_Y_GESTION_DE_BASE_DE DATOS=1 342	conf(1)	lift(1.03)	evr(0.02)	[8]		convr(4.83)

Figura 129. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante aprueba Diseño y Gestión de Base de Datos y Metodología de la programación es muy probable que también apruebe Metodología de la Investigación.
- 2) Si un estudiante aprueba Diseño y Gestión de Base de Datos, Metodología de la programación y Contabilidad General, es muy probable que también apruebe Metodología de la Investigación.



- 3) Si un estudiante es procedente de la ciudad de Loja y aprueba Diseño y Gestión de Base de Datos, es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 4) Si un estudiante es procedente del sector Urbano y aprueba Diseño y Gestión de Base de Datos y Metodología de la programación es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 5) Si un estudiante es procedente del sector Urbano y aprueba Diseño y Gestión de Base de Datos, Metodología de la programación y Contabilidad General es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 6) Si un estudiante efectúa alguna actividad laboral y aprueba la materia de Metodología de la programación es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 7) Si un estudiante efectúa alguna actividad laboral y aprueba Diseño y Gestión de Base de Datos y Metodología de la programación es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 8) Si un estudiante efectúa alguna actividad laboral y aprueba Metodología de la programación y Contabilidad General es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 9) Estudiante oriundo de la provincia de Loja y que apruebe Diseño y Gestión de Base de Datos y Contabilidad General es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 10) Si un estudiante procedente del sector Urbano, perteneciente a la ciudad de Loja y aprueba Diseño y Gestión de Base de Datos, es muy probable que también apruebe la asignatura de Metodología de la Investigación.
- 11) Si un estudiante ejerce alguna actividad laboral y aprueba Diseño y Gestión de Base de Datos, Metodología de la programación y Contabilidad General, es muy probable que también apruebe Metodología de la Investigación.
- 12) Si un estudiante es procedente del sector Urbano y aprueba Metodología de la Investigación, Metodología de la programación y Contabilidad

General, es muy probable que también apruebe la asignatura de Diseño y Gestión de Base de Datos.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado y con ello se han obtenido premisas con valores desde 1.14 en adelante, y en referencia se ha obtenido como máximo una confiabilidad del 82% en adelante (Ver Figura 4).

REGLAS	CONF	LIFT		
14. trabaja=0 tipo_colegio=2 METODOLOGIA_DE_LA_INVESTIGACION=1 170 ==> sector_procedencia=0 provincia_nacimiento=11 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 140 conf:	0.82	lift:(1.15)>	lev:(0.04)	[18]
15. trabaja=0 tipo_colegio=2 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 168 ==> sector_procedencia=0 provincia_nacimiento=11 METODOLOGIA_DE_LA_INVESTIGACION=1 140 conf:	0.83	lift:(1.15)>	lev:(0.04)	[18]
22. trabaja=0 tipo_colegio=2 METODOLOGIA_DE_LA_INVESTIGACION=1 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 167 ==> sector_procedencia=0 provincia_nacimiento=11 140 conf:	0.84	lift:(1.15)>	lev:(0.04)	[17]
29. trabaja=0 tipo_colegio=2 METODOLOGIA_DE_LA_INVESTIGACION=1 170 ==> sector_procedencia=0 provincia_nacimiento=11 142 conf:	0.84	lift:(1.14)>	lev:(0.04)	[17]
30. tipo_colegio=2 METODOLOGIA_DE_LA_INVESTIGACION=1 CONTABILIDAD_GENERAL=1 178 ==> sector_procedencia=0 provincia_nacimiento=11 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 145 conf:	0.81	lift:(1.14)>	lev:(0.04)	[17]
32. tipo_colegio=2 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 CONTABILIDAD_GENERAL=1 176 ==> sector_procedencia=0 provincia_nacimiento=11 METODOLOGIA_DE_LA_INVESTIGACION=1 145 conf:	0.82	lift:(1.14)>	lev:(0.04)	[17]
33. trabaja=0 tipo_colegio=2 174 ==> sector_procedencia=0 provincia_nacimiento=11 145 conf:	0.83	lift:(1.14)>	lev:(0.04)	[17]
34. trabaja=0 tipo_colegio=2 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 168 ==> sector_procedencia=0 provincia_nacimiento=11 140 conf:	0.83	lift:(1.14)>	lev:(0.04)	[17]
37. tipo_colegio=2 DISEÑO_Y_GESTION_DE_BASE_DE_DATOS=1 186 ==> sector_procedencia=0 provincia_nacimiento=11 METODOLOGIA_DE_LA_INVESTIGACION=1 153 conf:	0.82	lift:(1.14)>	lev:(0.04)	[18]

Figura 130. Reglas obtenidas en Weka con Lift mayor a 1

A continuación, se detallan los resultados obtenidos:

- 1) Si un estudiante no trabaja, ha culminado sus estudios en un colegio Fiscomisional, aprueba Metodología de la Investigación, es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Diseño y Gestión de Base de datos.
- 2) Si un estudiante no trabaja, ha culminado sus estudios en un colegio Fiscomisional, aprueba Diseño y Gestión de Base de datos, es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Metodología de la Investigación.
- 3) Si un estudiante no trabaja, ha culminado sus estudios secundarios en un colegio Fiscomisional, aprueba Metodología de la Investigación Diseño y Gestión de Base de datos, es muy probable que sea del sector Urbano y pertenezca a la provincia de Loja.
- 4) Si un estudiante no trabaja, ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Metodología de la Programación, es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Diseño y Gestión de Base de datos.
- 5) Si un estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Metodología de la Programación y Contabilidad General es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Diseño y Gestión de Base de datos.

- 6) Si un estudiante ha culminado sus estudios secundarios en un colegio Fiscomisional y aprueba Diseño y Gestión de Base de Datos y Contabilidad General es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Metodología de la Programación.
- 7) Si un estudiante no trabaja, ha culminado sus estudios en un colegio Fiscomisional, es muy probable que sea del sector Urbano y que pertenezca a la provincia de Loja.
- 8) Si un estudiante no trabaja, ha culminado sus estudios en un colegio Fiscomisional y aprueba Diseño y Gestión de Base de datos, es muy probable que sea del sector Urbano y que pertenezca a la provincia de Loja.
- 9) Si un estudiante ha culminado sus estudios en un colegio Fiscomisional y aprueba Diseño y Gestión de Base de datos, es muy probable que sea del sector Urbano, pertenezca a la provincia de Loja y que apruebe Metodología de la Programación.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional, de formar parte del sector Urbano y de ser procedente de la provincia de Loja, están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es muy baja, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, sin embargo, en este caso se muestra una variable más, que es el tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a



partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 22 a 26 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

#### Association Rules

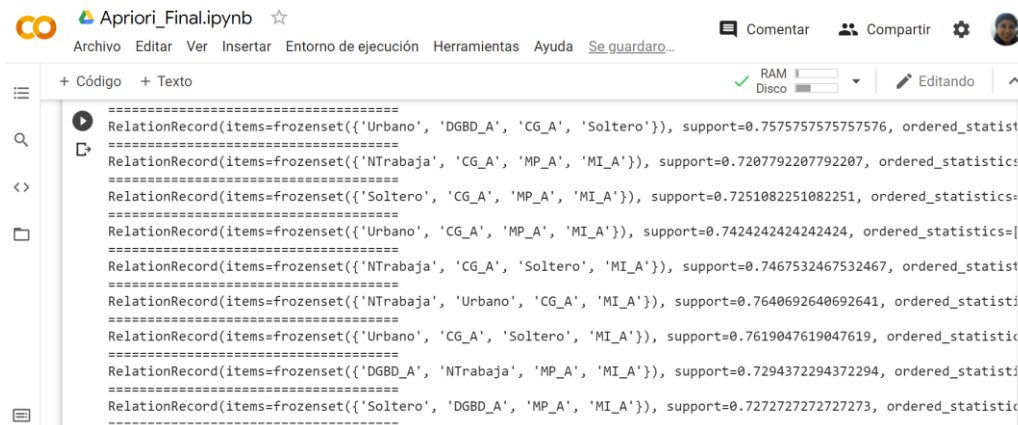
```
[tipo_colegio] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.801)
[tipo_colegio] --> [METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.801)
[tipo_colegio] --> [METODOLOGÍA DE LA INVESTIGACIÓN, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.801)
[edad, tipo_colegio] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.802)
[edad, tipo_colegio] --> [CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.802)
[edad, tipo_colegio] --> [METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.804)
[edad, tipo_colegio] --> [METODOLOGÍA DE LA INVESTIGACIÓN, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.802)
[edad] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.804)
[edad] --> [METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.804)
[genero] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.805)
[genero] --> [METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.806)
[tipo_colegio] --> [CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.806)
[METODOLOGÍA DE LA INVESTIGACIÓN, edad] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.81)
[edad] --> [METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.808)
[edad] --> [METODOLOGÍA DE LA INVESTIGACIÓN, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.808)
[DISEÑO Y GESTIÓN DE BASE DE DATOS, edad, tipo_colegio] --> [CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.810)
[DISEÑO Y GESTIÓN DE BASE DE DATOS, edad, tipo_colegio] --> [METODOLOGÍA DE LA INVESTIGACIÓN, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.810)
[METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, edad, tipo_colegio] --> [CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.810)
[METODOLOGÍA DE LA INVESTIGACIÓN, genero] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[METODOLOGÍA DE LA INVESTIGACIÓN, tipo_colegio] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[DISEÑO Y GESTIÓN DE LA INVESTIGACIÓN] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[DISEÑO Y GESTIÓN DE BASE DE DATOS, edad] --> [METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[DISEÑO Y GESTIÓN DE BASE DE DATOS, edad] --> [METODOLOGÍA DE LA INVESTIGACIÓN, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[METODOLOGÍA DE LA INVESTIGACIÓN, DISEÑO Y GESTIÓN DE BASE DE DATOS, edad] --> [METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.811)
[METODOLOGÍA DE LA INVESTIGACIÓN, edad, tipo_colegio] --> [DISEÑO Y GESTIÓN DE BASE DE DATOS, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.812)
[METODOLOGÍA DE LA INVESTIGACIÓN, edad, tipo_colegio] --> [CONTABILIDAD GENERAL, METODOLOGÍA DE LA PROGRAMACIÓN] (confidence: 0.812)
```

Figura 131. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que sean aprobadas las materias el estudiante dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, si el estado civil del alumno es soltero también aprueban todas las materias, si los estudiantes pertenecen al sector Urbano y son oriundo de la provincia de Loja y finalmente, con la aprobación de algunas materias relacionadas de forma paralela, se muestra relación entre las asignaturas Metodología de la Programación, Metodología de la Investigación y Diseño y Gestión de Base de Datos, algunas en conjunto con otras características necesarias para la aprobación de las mismas, además, se

mantiene la relación con la materia de Contabilidad General, ésta se la aprueba mientras se la lleve en ejecución de las materias descritas anteriormente.



```

=====
RelationRecord(items=frozenset({'Urbano', 'DGBD_A', 'CG_A', 'Soltero'}), support=0.7575757575757576, ordered_statistic=0.7575757575757576)
=====
RelationRecord(items=frozenset({'NTrabaja', 'CG_A', 'MP_A', 'MI_A'}), support=0.7207792207792207, ordered_statistic=0.7207792207792207)
=====
RelationRecord(items=frozenset({'Soltero', 'CG_A', 'MP_A', 'MI_A'}), support=0.7251082251082251, ordered_statistic=0.7251082251082251)
=====
RelationRecord(items=frozenset({'Urbano', 'CG_A', 'MP_A', 'MI_A'}), support=0.7424242424242424, ordered_statistic=0.7424242424242424)
=====
RelationRecord(items=frozenset({'NTrabaja', 'CG_A', 'Soltero', 'MI_A'}), support=0.7467532467532467, ordered_statistic=0.7467532467532467)
=====
RelationRecord(items=frozenset({'NTrabaja', 'Urbano', 'CG_A', 'MI_A'}), support=0.7640692640692641, ordered_statistic=0.7640692640692641)
=====
RelationRecord(items=frozenset({'Urbano', 'CG_A', 'Soltero', 'MI_A'}), support=0.7619047619047619, ordered_statistic=0.7619047619047619)
=====
RelationRecord(items=frozenset({'DGBD_A', 'NTrabaja', 'MP_A', 'MI_A'}), support=0.7294372294372294, ordered_statistic=0.7294372294372294)
=====
RelationRecord(items=frozenset({'Soltero', 'DGBD_A', 'MP_A', 'MI_A'}), support=0.7272727272727273, ordered_statistic=0.7272727272727273)
=====

```

Figura 132. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances		
0	205	( 44%)
1	31	( 7%)
2	174	( 38%)
3	26	( 6%)
4	26	( 6%)

Figura 133. Clúster con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Cluster#					
	Full Data (462.0)	0 (205.0)	1 (31.0)	2 (174.0)	3 (26.0)	4 (26.0)
edad	29	28	22	25	27	26
genero	0	0	0	0	0	0
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	7
trabaja	0	0	0	0	0	0
tipo_colegio	1	2	2	1	1	1
METODOLOGIA_DE_LA_INVESTIGACION	1	1	1	1	1	1
DISEÑO_Y_GESTION_DE_BASE_DE_DATOS	1	1	1	1	1	1
METODOLOGIA_DE_LA_PROGRAMACION	1	1	1	1	1	1
CONTABILIDAD_GENERAL	1	1	1	1	1	1

Figura 134. Clúster obtenidos por Weka.

**Clúster\_0 (44%):** Estudiantes que tengan una edad de 28 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, es decir, Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General.

**Clúster\_1 (7%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, es decir, Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General.

**Clúster\_2 (38%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, es decir, Metodología de la Investigación,



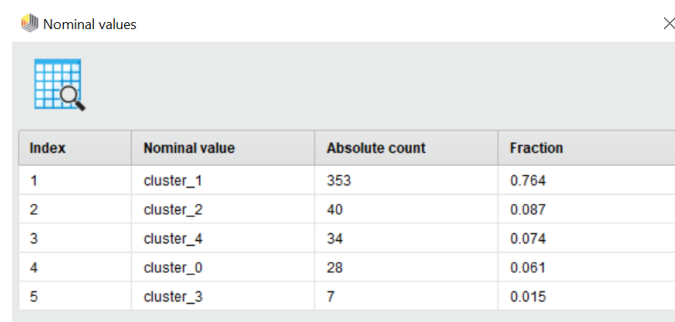
Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General.

**Clúster\_3 (6%):** Estudiantes que tengan una edad de 27 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, es decir, Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General.

**Clúster\_4 (6%):** Estudiantes que tengan una edad de 26 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de El Oro, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, es decir, Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_1	353	0.764
2	cluster_2	40	0.087
3	cluster_4	34	0.074
4	cluster_0	28	0.061
5	cluster_3	7	0.015

Figura 135. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 671 instancias están clasificadas o agrupadas a su clúster más cercano.

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	METODOLO...	DISEÑO Y GE...	MEI
1	1	cluster_4	-0.306	1.571	3.489	-0.127	-0.332	0.889	0.124	-5.871	0.45
2	2	cluster_1	0.517	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	-2.1
3	3	cluster_1	1.614	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	0.45
4	4	cluster_0	-0.031	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	-2.1
5	5	cluster_1	-1.403	-0.635	-0.286	2.472	-0.332	-0.867	0.124	0.170	0.45
6	6	cluster_1	-0.580	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	0.45
7	7	cluster_1	-0.031	-0.635	-0.286	-1.427	-0.332	-0.867	0.124	0.170	0.45
8	8	cluster_1	0.243	-0.635	-0.286	-1.427	-0.332	-0.867	0.124	0.170	0.45
9	9	cluster_1	-0.580	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	0.45
10	10	cluster_1	-0.031	-0.635	-0.286	-0.127	-0.332	-0.867	0.124	0.170	0.45
11	11	cluster_1	-1.129	-0.635	-0.286	-0.127	-0.332	0.889	0.124	0.170	-2.1
12	12	cluster_1	0.243	-0.635	-0.286	-0.127	-0.332	0.889	0.124	0.170	0.45
13	13	cluster_1	-1.403	-0.635	-0.286	-0.127	-0.332	0.889	0.124	0.170	0.45

Figura 136. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 1 con más instancias agrupadas.

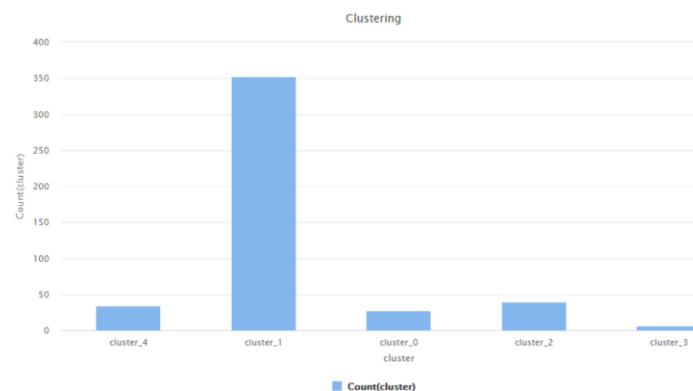


Figura 137. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

### k-Means - Centroid Table

Cluster	CONTABIL...	DISENO Y ...	edad	genero	METODOL...	METODOL...	provincia_...	sector_pr...	tipo_colegio	trabaja
Cluster 0	1	1	26.355	0.020	1	0.868	11.493	0.072	2.105	0.099
Cluster 1	1	1	27.436	1	1	0.889	11.521	0.043	1.333	0.051
Cluster 2	0.500	0.071	24.571	0.357	0.500	0.429	11.857	0.143	1.786	0.143
Cluster 3	1	1	27.412	0	1	0.869	11.307	0.105	1	0.144
Cluster 4	0.000	1	29.731	0.308	1	0.269	10.462	0.038	1.385	0.038

Figura 138. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas tales como Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General. todas pertenecientes al cuarto ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas tales como Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General. todas pertenecientes al cuarto ciclo, si tiene una edad de 27 años, es de género femenino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

**Clúster 2:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas tales como Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General. todas pertenecientes al cuarto ciclo, si tiene una edad de 24 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

**Clúster 3:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas tales como Metodología de la Investigación, Diseño y Gestión de Base de datos, Metodología de la programación y Contabilidad General. todas pertenecientes al cuarto ciclo, si tiene una edad de 27 años, es de género



masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

**Clúster 4:** Estudiante aprueba únicamente a las asignaturas de Diseño y Gestión de Base de datos, Metodología de la programación, pertenecientes al cuarto ciclo, si tiene una edad de 29 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no ejerce ninguna actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.

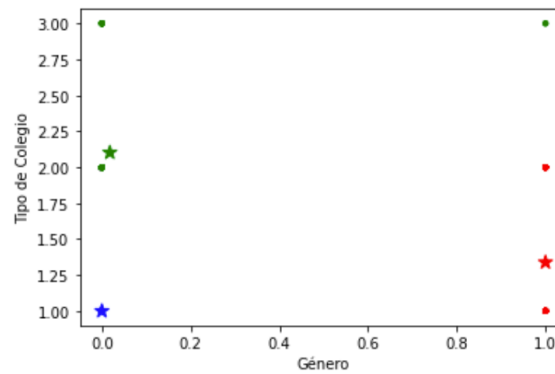


Figura 139. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Metodología de la Programación**, en donde los estudiantes aprueban a la asignatura si su edad es de 23, 28 y 33 años.

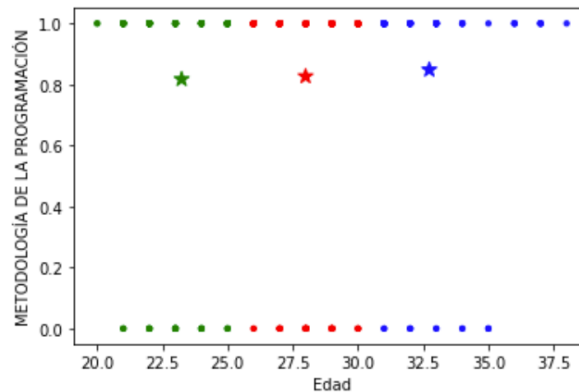


Figura 140. Relación Edad – Metodología de la Programación

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Metodología de la Programación**, donde se observa que un grupo mínimo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que probabilísticamente tiendan a reprobado la materia de Metodología de la Programación.

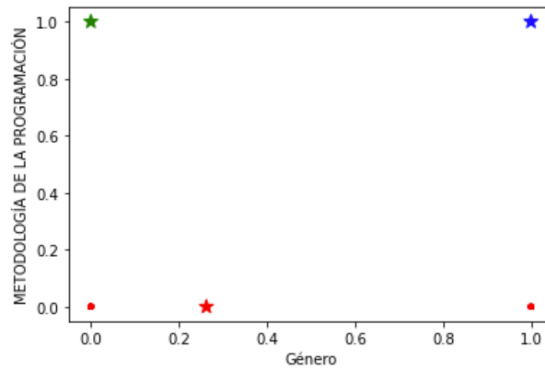


Figura 141. Relación Género – Metodología de la Programación

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Diseño y Gestión de Base de Datos** y **Metodología de la Programación**, en un clúster (rojo) se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso contrario, en donde no se aprueba ninguna de las dos asignaturas y, por último, se da el caso donde aprobará Diseño y Gestión de Base de Datos y no Metodología de la Programación.

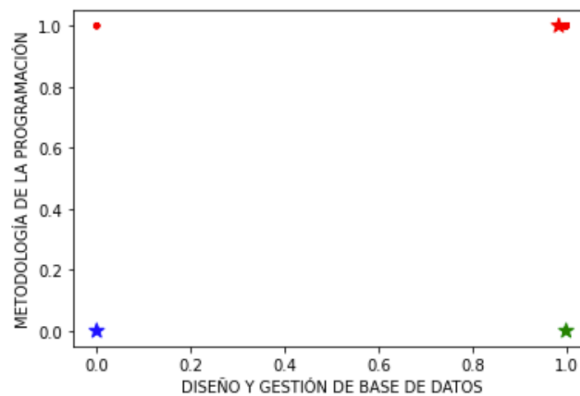


Figura 142. Relación Diseño y Gestión de Base de Datos – Metodología de la Programación

## Anexo 8. Ingeniería en Sistemas – Quinto ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **quinto ciclo de la carrera de Ingeniería en Sistemas** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

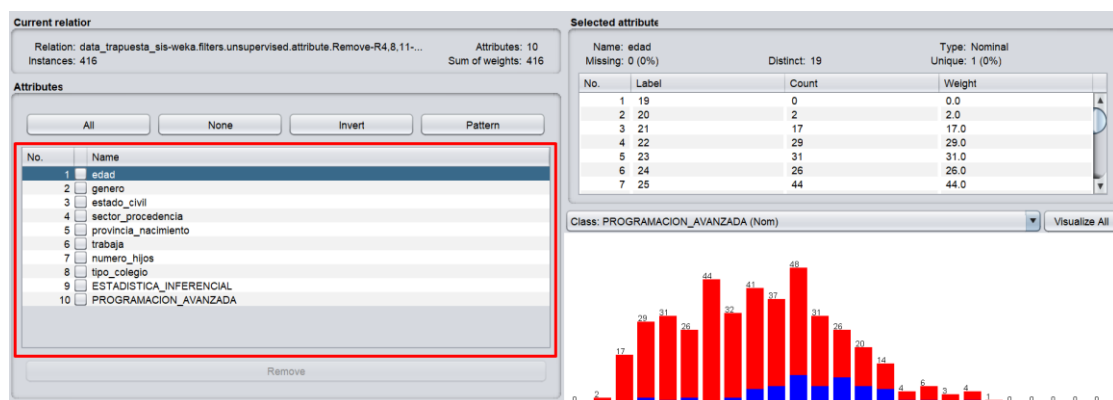


Figura 143. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.









- Estadística Inferencial es muy probable que apruebe Programación Avanzada.
- 3) Si un estudiante es de género masculino, no ejerce una actividad Laboral, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
  - 4) Si un estudiante es de género masculino, procedente del sector Urbano, no ejerce actividad laboral, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
  - 5) Si un estudiante es de género masculino, pertenece a la provincia de Loja, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
  - 6) Si un estudiante es de género masculino, procedente del sector Urbano, pertenece a la provincia de Loja, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
  - 7) Si un estudiante aprueba la asignatura de Estadística Inferencial es muy probable que se dé debido a que no ejerce alguna actividad laboral.
  - 8) Si un estudiante es de género masculino, pertenece a la provincia de Loja, culmina sus estudios secundarios en una Institución Fiscal y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
  - 9) Si un estudiante es de género femenino y culmina sus estudios secundarios en una Institución Fiscal es muy probable que no ejerza ninguna actividad laboral.
  - 10) Si un estudiante es de género masculino y aprueba Estadística Inferencial es muy probable que también apruebe Programación Avanzada.
  - 11) Si un estudiante es de género femenino, procedente del sector Urbano y culmina sus estudios secundarios en una Institución Fiscal es muy probable que no ejerza alguna actividad laboral.
  - 12) Si un estudiante aprueba la asignatura de Estadística Inferencial es muy probable que se dé debido a que no ejerce alguna actividad laboral.

- 13) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad laboral, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba Programación Avanzada es muy probable que se dé debido a que pertenece al sector Urbano.
- 14) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad laboral y culmina sus estudios secundarios en una Institución Fiscomisional es muy probable que se dé debido a que pertenece al sector Urbano.
- 15) Si un estudiante es de género masculino, procedente del sector Urbano y aprueba Estadística Inferencial es muy probable que apruebe Programación Avanzada.
- 16) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad laboral, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba Estadística Inferencial es muy probable que este pertenezca al sector Urbano.
- 17) Si un estudiante es de género femenino, culmina sus estudios secundarios en una Institución Fiscal y aprueba la asignatura de Estadística Inferencial es muy probable que sea debido a que el estudiante no ejerce actividad laboral.
- 18) Si un estudiante no ejerce actividad laboral, y culmina sus estudios secundarios en una Institución Fiscomisional es muy probable que se dé debido a que el estudiante es procedente del sector Urbano.

Por otra parte, se han determinado Reglas definiendo Lift como predeterminado, en este caso con valores mayores a 1.21 pero las premisas muestran una confiabilidad baja (30%), por ello se ha filtrado la confianza para obtener únicamente las reglas con un lift elevado y una confianza mayor al 80% (Ver Figura 4). A continuación, se muestran los resultados obtenidos:

REGLAS	CONF.	LIFT	ev:	[	conv:
90. genero=0 tipo_colegio=2 PROGRAMACION_AVANZADA=1 115 ==> provincia_nacimiento=11 ESTADISTICA_INFERENCIAL=1 92 conf:	0.8)	lift:(1.23)>	ev:(0.04)	[17]	conv:(1.67)
99. genero=0 sector_procedencia=0 tipo_colegio=2 PROGRAMACION_AVANZADA=1 109 ==> provincia_nacimiento=11 ESTADISTICA_INFERENCIAL=1 87	0.8)	lift:(1.23)>	ev:(0.04)	[15]	conv:(1.65)
117. genero=0 tipo_colegio=2 ESTADISTICA_INFERENCIAL=1 114 ==> provincia_nacimiento=11 PROGRAMACION_AVANZADA=1 92 conf:	0.81)	lift:(1.21)>	ev:(0.04)	[16]	conv:(1.66)
121. genero=0 sector_procedencia=0 tipo_colegio=2 ESTADISTICA_INFERENCIAL=1 108 ==> provincia_nacimiento=11 PROGRAMACION_AVANZADA=1 87	0.81)	lift:(1.21)>	ev:(0.04)	[15]	conv:(1.64)

Figura 146. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante es de género masculino, ha culminado sus estudios en un colegio Fiscomisional y aprueba Programación Avanzada, es muy

- probable que pertenezca a la provincia de Loja y que apruebe Estadística Inferencial.
- 2) Si un estudiante es de género masculino, procedente del sector Urbano, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la asignatura de Programación Avanzada es muy probable que pertenezca a la provincia de Loja y que apruebe Estadística Inferencial.
  - 3) Si un estudiante es de género masculino, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la asignatura de Programación Avanzada es muy probable que pertenezca a la provincia de Loja y que apruebe Estadística Inferencial.
  - 4) Si un estudiante es de género masculino, procedente del sector Urbano, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la asignatura de Estadística Inferencial es muy probable que pertenezca a la provincia de Loja y que apruebe Programación Avanzada.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular mínima, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo, además de que, como se refleja en las premisas se relaciona más con variables demográficas y sociales que con las académicas, considerando específicamente a que no se muestra una relación con la lógica que se necesita para la asignatura de Programación Avanzada.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las únicas asignaturas relacionadas a las matemáticas se mantiene, a estas se suman variables como el tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que

los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 23 a 28 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

Association Rules

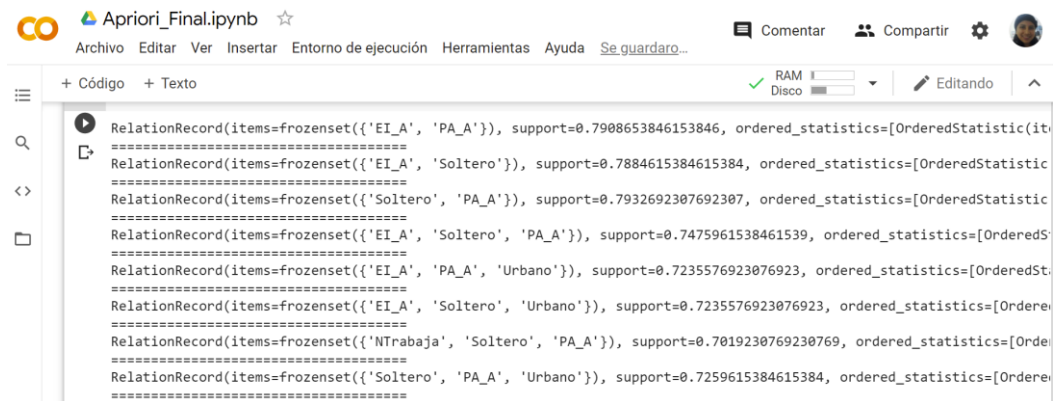
```
[PROGRAMACION AVANZADA, edad, tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.780)
[edad, tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.792)
[tipo_colegio, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.814)
[PROGRAMACION AVANZADA, edad] --> [ESTADISTICA INFERENCIAL] (confidence: 0.826)
[genero] --> [PROGRAMACION AVANZADA] (confidence: 0.829)
[genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.829)
[PROGRAMACION AVANZADA, tipo_colegio, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.833)
[PROGRAMACION AVANZADA, edad, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.836)
[tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.837)
[tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.846)
[tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.856)
[ESTADISTICA INFERENCIAL, tipo_colegio, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.857)
[ESTADISTICA INFERENCIAL, edad, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.868)
[sector_procedencia] --> [ESTADISTICA INFERENCIAL] (confidence: 0.875)
[sector_procedencia] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.875)
[ESTADISTICA INFERENCIAL, edad, tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.885)
[PROGRAMACION AVANZADA, genero] --> [ESTADISTICA INFERENCIAL] (confidence: 0.892)
[ESTADISTICA INFERENCIAL, genero] --> [PROGRAMACION AVANZADA] (confidence: 0.892)
[sector_procedencia] --> [PROGRAMACION AVANZADA] (confidence: 0.906)
[PROGRAMACION AVANZADA, tipo_colegio] --> [ESTADISTICA INFERENCIAL] (confidence: 0.907)
[ESTADISTICA INFERENCIAL, edad] --> [PROGRAMACION AVANZADA] (confidence: 0.911)
[PROGRAMACION AVANZADA] --> [ESTADISTICA INFERENCIAL] (confidence: 0.916)
[ESTADISTICA INFERENCIAL, tipo_colegio] --> [PROGRAMACION AVANZADA] (confidence: 0.918)
[trabaja] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.932)
[ESTADISTICA INFERENCIAL] --> [PROGRAMACION AVANZADA] (confidence: 0.940)
[provincia_nacimiento] --> [PROGRAMACION AVANZADA, ESTADISTICA INFERENCIAL] (confidence: 0.953)
[trabaja] --> [PROGRAMACION AVANZADA] (confidence: 0.955)
[trabaja] --> [ESTADISTICA INFERENCIAL] (confidence: 0.955)
```

Figura 147. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que sean aprobadas las materias el estudiante dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, si el estado civil del alumno es soltero también aprueban todas las materias, si los estudiantes pertenecen al

sector Urbano y finalmente, en este ciclo se da únicamente dos materias relacionadas a las matemáticas Estadística Inferencial y Programación avanzada y si bien es cierto las dos se encuentran vinculadas junto con los atributos antes mencionados para su respectiva aprobación.



```

RelationRecord(items=frozenset({'EI_A', 'PA_A'}), support=0.7908653846153846, ordered_statistics=[OrderedStatistic(it
=====
RelationRecord(items=frozenset({'EI_A', 'Soltero'}), support=0.7884615384615384, ordered_statistics=[OrderedStatistic
=====
RelationRecord(items=frozenset({'Soltero', 'PA_A'}), support=0.7932692307692307, ordered_statistics=[OrderedStatistic
=====
RelationRecord(items=frozenset({'EI_A', 'Soltero', 'PA_A'}), support=0.7475961538461539, ordered_statistics=[OrderedS
=====
RelationRecord(items=frozenset({'EI_A', 'PA_A', 'Urbano'}), support=0.7235576923076923, ordered_statistics=[OrderedSt
=====
RelationRecord(items=frozenset({'EI_A', 'Soltero', 'Urbano'}), support=0.7235576923076923, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'NTrabaja', 'Soltero', 'PA_A'}), support=0.7019230769230769, ordered_statistics=[Orde
=====
RelationRecord(items=frozenset({'Soltero', 'PA_A', 'Urbano'}), support=0.7259615384615384, ordered_statistics=[Order
=====

```

Figura 148. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances		
0	236	( 57%)
1	83	( 20%)
2	33	( 8%)
3	27	( 6%)
4	37	( 9%)

Figura 149. clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Cluster#					
	Full Data (416.0)	0 (236.0)	1 (83.0)	2 (33.0)	3 (27.0)	4 (37.0)
edad	29	25	28	23	22	29
genero	0	0	1	0	0	1
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	1	0	0
tipo_colegio	1	2	1	1	2	1
ESTADISTICA_INFERENCIAL	1	1	1	1	1	1
PROGRAMACION_AVANZADA	1	1	1	1	1	1

Figura 150. Clúster obtenidos por Weka.

**Clúster\_0 (57%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_1 (20%):** Estudiantes que tengan una edad de 28 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_2 (8%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, cuenta con una actividad laboral y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_3 (6%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución

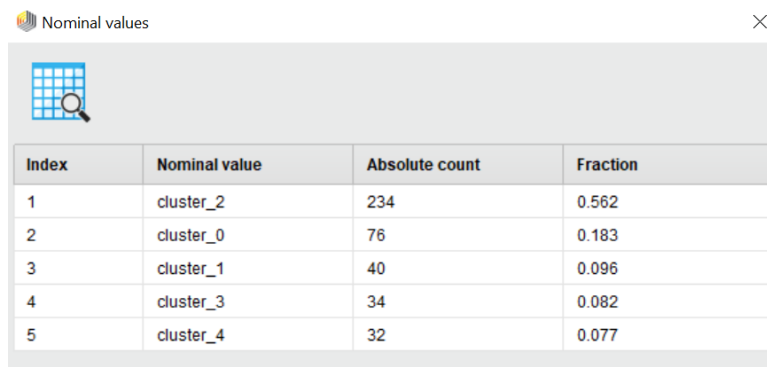


Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

**Clúster\_4 (9%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al quinto ciclo, estas son: Estadística Inferencial y Programación Avanzada.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_2	234	0.562
2	cluster_0	76	0.183
3	cluster_1	40	0.096
4	cluster_3	34	0.082
5	cluster_4	32	0.077

Figura 151. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 416 instancias están clasificadas o agrupadas a su clúster más cercano.

Result History: ExampleSet (Clustering) | PerformanceVector (Performance) | ExampleSet (/Local Repository/sis\_03)

Open in: Turbo Prep | Auto Model | Filter (416 / 416 examples): all

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ESTADISTIC...	PROGRAMA...
1	1	cluster_4	-0.327	1.542	3.460	-0.125	-0.344	0.912	0.434	0.398
2	2	cluster_2	1.574	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
3	3	cluster_3	-1.413	-0.647	-0.288	2.500	-0.344	-0.861	0.434	0.398
4	4	cluster_2	-0.599	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
5	5	cluster_2	-0.055	-0.647	-0.288	-1.438	-0.344	-0.861	0.434	0.398
6	6	cluster_2	0.216	-0.647	-0.288	-1.438	-0.344	-0.861	0.434	0.398
7	7	cluster_2	-0.599	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
8	8	cluster_2	-0.055	-0.647	-0.288	-0.125	-0.344	-0.861	0.434	0.398
9	9	cluster_2	-1.142	-0.647	-0.288	-0.125	-0.344	0.912	0.434	0.398
10	10	cluster_0	0.488	1.542	-0.288	-0.125	-0.344	0.912	-2.300	-2.507
11	11	cluster_2	-1.413	-0.647	-0.288	-0.125	-0.344	0.912	0.434	0.398
12	12	cluster_2	-1.685	1.542	-0.288	-0.125	-0.344	-0.861	0.434	0.398
13	13	cluster_3	-0.599	-0.647	-0.288	1.844	-0.344	0.912	0.434	0.398

Figura 152. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 2 con más instancias agrupadas.

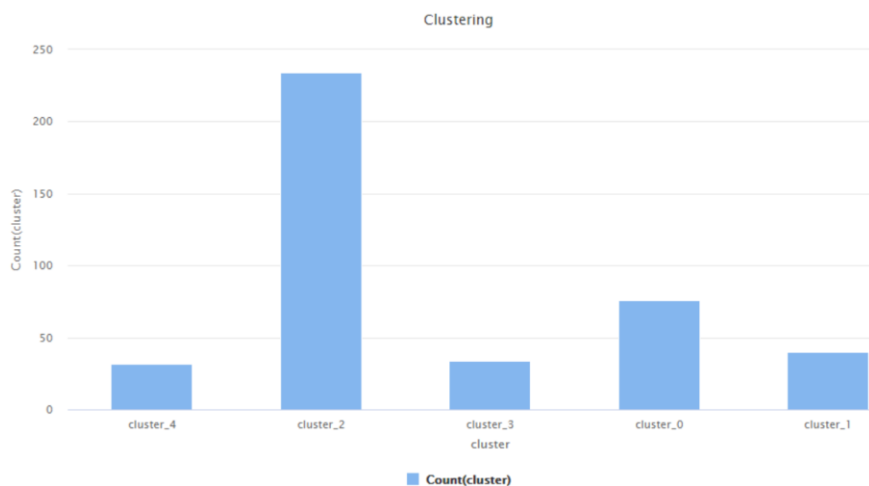


Figura 153. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.



### k-Means - Centroid Table

Cluster	edad	ESTADISTICA L...	genero	PROGRAMACI...	provincia_nac...	sector_proce...	tipo_colegio	trabaja
Cluster 0	26.634	0.976	0.146	0.976	11.927	0	1.439	1
Cluster 1	27.094	0.875	0.219	0.906	11.625	1.000	1.469	0.094
Cluster 2	27.359	0.978	1	0.891	11.772	0	1.370	-0
Cluster 3	26.173	1	0	0.953	11.277	-0	1.565	-0
Cluster 4	30.700	0.017	0.300	0.433	10.617	0	1.450	-0

Figura 154. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 26 años, es de género masculino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscomisional y que cuenta con una actividad laboral extra a partir de sus actividades académicas.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 27 años, es de género masculino, perteneciente a la provincia de Loja y al sector Rural, proviene de un colegio Fiscal y que no cuenta con una actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 27 años, es de género femenino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscal y que no cuenta con una actividad laboral.

**Clúster 3:** Estudiante aprueba únicamente la asignatura de Geometría plana y reprueba Cálculo Diferencial y Fundamentos Informáticos, todas pertenecientes al primer ciclo, si tiene una edad de 26 años, es soltero, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 0:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como Estadística Inferencial y Programación Avanzada, todas pertenecientes al quinto ciclo, si este tiene una edad de 30 años, es de género

masculino, perteneciente a la provincia de Loja y al sector Urbano, proviene de un colegio Fiscal y no cuenta con una actividad laboral.

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, con ello se logra identificar que no hay muchos estudiantes que hayan culminado sus estudios en una Institución Particular.

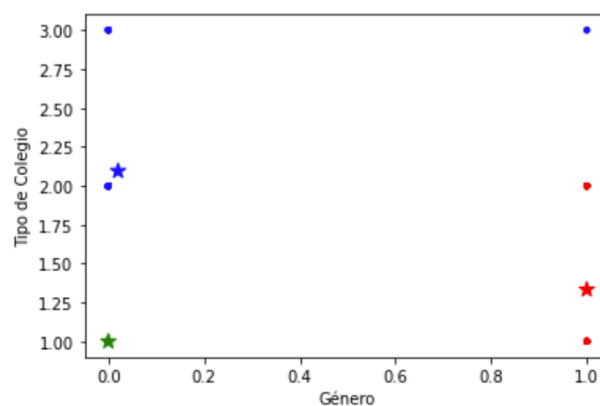


Figura 155. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Metodología de la Programación**, en donde los estudiantes aprueban a la asignatura si su edad es de 24, 28 y 33 años. Cabe recalcar que estudiantes con edad de 33 años, pueden probabilísticamente perder a dicha asignatura.

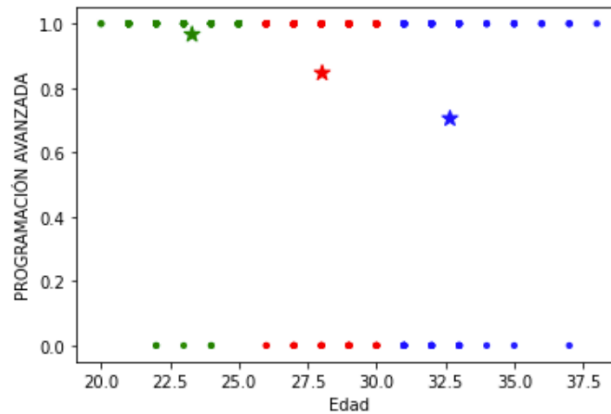


Figura 156. Relación Edad – Metodología de la Programación

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Programación Avanzada**, donde se observa que un grupo mínimo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género con probabilidad de reprobar la materia de Programación Avanzada.

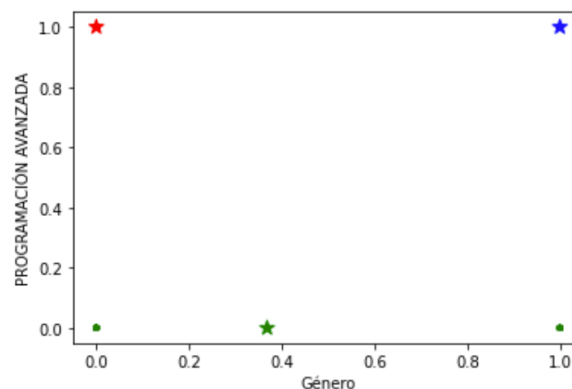


Figura 157. Relación Género – Metodología de la Programación

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Estadística Inferencial** y **Programación Avanzada**, en un clúster (verde) se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde muy probable se apruebe a Estadística Inferencial pero reprobará Programación avanzada, mientras que aprobará Programación Avanzada pero no tendrá ninguna posibilidad de aprobar Estadística inferencial.

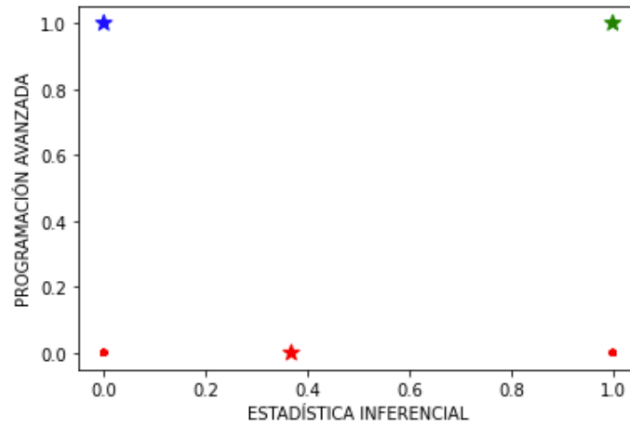


Figura 158. Relación Diseño y Gestión de Base de Datos – Metodología de la Programación

## Anexo 9. Ingeniería en Sistemas – Sexto ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **sexto ciclo de la carrera de Ingeniería en Sistemas** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

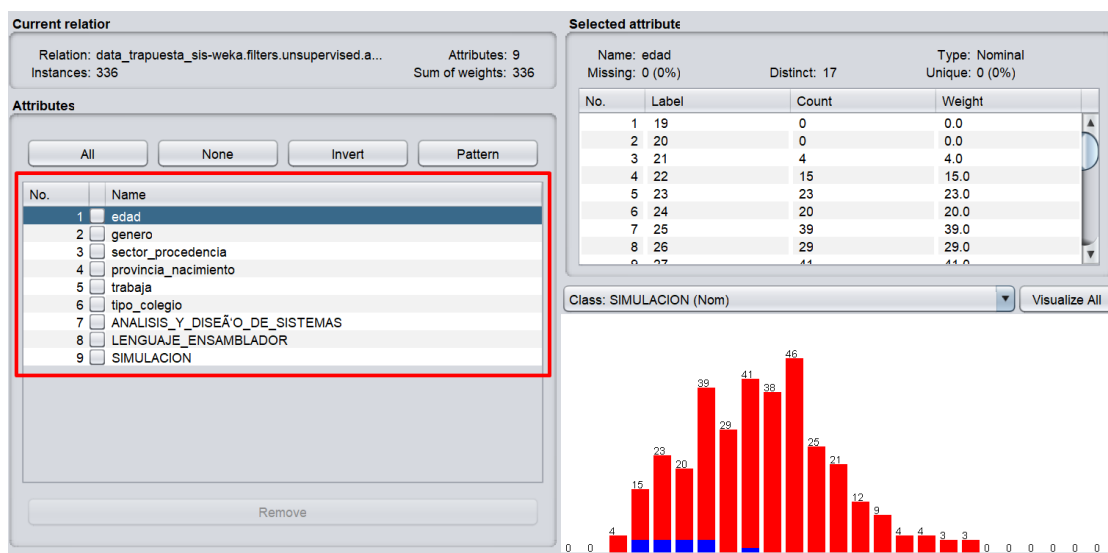


Figura 159. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. trabaja=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 282 ==> sector_procedencia=0 SIMULACION=1 258 <conf:(0.91)> lift:(1.03)
177. ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 316 ==> sector_procedencia=0 SIMULACION=1 289 <conf:(0.91)> lift:(1.03)
178. genero=0 233 ==> sector_procedencia=0 213 <conf:(0.91)> lift:(0.99) lev:(-0.01) [-1] conv:(0.86)
179. genero=0 233 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 213 <conf:(0.91)> lift:(0.99) lev:(-0.01)
180. provincia_nacimiento=11 trabaja=0 230 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 210 <conf:(0.91)> lift:(1.04)
181. provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 241 ==> sector_procedencia=0 LENGUAJE_ENSAMBLADOR=1 210 <conf:(0.91)> lift:(1.04)
182. sector_procedencia=0 provincia_nacimiento=11 241 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 210 <conf:(0.91)> lift:(1.04)
183. trabaja=0 298 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 272 <conf:(0.91)> lift:(1) lev:(0.02)
184. LENGUAJE_ENSAMBLADOR=1 318 ==> sector_procedencia=0 SIMULACION=1 290 <conf:(0.91)> lift:(1.02) lev:(0.02)
185. provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 246 ==> sector_procedencia=0 SIMULACION=1 224 <conf:(0.91)> lift:(1.04)
186. LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 312 ==> sector_procedencia=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 284 <conf:(0.91)> lift:(1.04)
187. trabaja=0 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 278 ==> sector_procedencia=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 284 <conf:(0.91)> lift:(1.04)
188. genero=0 233 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 212 <conf:(0.91)> lift:(1) lev:(0.02)
189. provincia_nacimiento=11 262 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 238 <conf:(0.91)> lift:(1.04)
190. provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 248 ==> sector_procedencia=0 SIMULACION=1 225 <conf:(0.91)> lift:(1.04)
191. trabaja=0 298 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 270 <conf:(0.91)> lift:(1.04)
192. provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 243 ==> sector_procedencia=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 284 <conf:(0.91)> lift:(1.04)
193. ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 316 ==> sector_procedencia=0 LENGUAJE_ENSAMBLADOR=1 286 <conf:(0.91)> lift:(1.04)
194. provincia_nacimiento=11 trabaja=0 230 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 278 <conf:(0.91)> lift:(1.04)
195. trabaja=0 LENGUAJE_ENSAMBLADOR=1 282 ==> sector_procedencia=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 255 <conf:(0.91)> lift:(1.04)
196. trabaja=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 282 ==> sector_procedencia=0 LENGUAJE_ENSAMBLADOR=1 255 <conf:(0.91)> lift:(1.04)
197. provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 246 ==> sector_procedencia=0 LENGUAJE_ENSAMBLADOR=1 210 <conf:(0.91)> lift:(1.04)
198. genero=0 233 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 210 <conf:(0.91)> lift:(1.04)
199. provincia_nacimiento=11 262 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 236 <conf:(0.91)> lift:(1.04)
200. provincia_nacimiento=11 trabaja=0 230 ==> sector_procedencia=0 SIMULACION=1 207 <conf:(0.91)> lift:(1.04)
  
```

Figura 160. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza con un 97% de confiabilidad en las premisas obtenidas y un lift con valores mayores a 1.04. Con ello se obtiene un total de 12 reglas (Ver Figura 3).

REGLAS	CONF.	LIFT.	LEV.	CONV.
11. genero=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 213 ==> LENGUAJE_ENSAMBLADOR=1 210	conf:(0.99)	lift:(1.04)	lev:(0.03)	[8]
14. sector_procedencia=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 289 ==> LENGUAJE_ENSAMBLADOR=1 284	conf:(0.98)	lift:(1.04)	lev:(0.03)	[10]
16. sector_procedencia=0 provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 224 ==> LENGUAJE_ENSAMBLADOR=1 220	conf:(0.98)	lift:(1.04)	lev:(0.02)	[8]
20. ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 310 ==> LENGUAJE_ENSAMBLADOR=1 304	conf:(0.98)	lift:(1.04)	lev:(0.03)	[10]
21. sector_procedencia=0 trabaja=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 258 ==> LENGUAJE_ENSAMBLADOR=1 253	conf:(0.98)	lift:(1.04)	lev:(0.03)	[8]
25. sector_procedencia=0 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 290 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 284	conf:(0.98)	lift:(1.04)	lev:(0.03)	[11]
30. sector_procedencia=0 provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 225 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 220	conf:(0.98)	lift:(1.04)	lev:(0.02)	[8]
34. sector_procedencia=0 trabaja=0 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 259 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 253	conf:(0.98)	lift:(1.04)	lev:(0.03)	[9]
35. genero=0 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 215 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 210	conf:(0.98)	lift:(1.04)	lev:(0.02)	[7]
39. LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 312 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 304	conf:(0.97)	lift:(1.04)	lev:(0.03)	[10]
53. genero=0 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 217 ==> LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 210	conf:(0.97)	lift:(1.04)	lev:(0.03)	[8]
60. sector_procedencia=0 trabaja=0 LENGUAJE_ENSAMBLADOR=1 262 ==> ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 253	conf:(0.97)	lift:(1.05)	lev:(0.03)	[11]

Figura 161. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:



- 1) Si un estudiante de género masculino, aprueba las materias de Análisis y Diseño de Sistemas y Simulación es muy probable que también apruebe la asignatura de Lenguaje Ensamblador.
- 2) Si un estudiante es procedente al sector Urbano, aprueba las materias de Análisis y Diseño de Sistemas y Simulación es muy probable que también apruebe la asignatura de Lenguaje Ensamblador.
- 3) Si un estudiante es procedente al sector Urbano, pertenece a la Provincia de Loja y aprueba las materias de Análisis y Diseño de Sistemas y Simulación es muy probable que también apruebe la asignatura de Lenguaje Ensamblador.
- 4) Si un estudiante aprueba las materias Análisis y Diseño de Sistemas y Simulación es muy probable que también apruebe la asignatura de Lenguaje Ensamblador.
- 5) Si un estudiante es procedente al sector Urbano, no ejerce actividad laboral y aprueba las materias de Análisis y Diseño de Sistemas y Simulación es muy probable que también apruebe la asignatura de Lenguaje Ensamblador.
- 6) Si un estudiante es procedente al sector Urbano y aprueba las materias de Lenguaje Ensamblador y Simulación es muy probable que también apruebe la asignatura de Análisis y Diseño de Sistemas.
- 7) Si un estudiante es procedente al sector Urbano, pertenece a la provincia de Loja y aprueba las materias de Lenguaje Ensamblador y Simulación es muy probable que también apruebe la asignatura de Análisis y Diseño de Sistemas.
- 8) Si un estudiante es procedente al sector Urbano, no ejerce actividad laboral y aprueba las materias de Lenguaje Ensamblador y Simulación es muy probable que también apruebe la asignatura de Análisis y Diseño de Sistemas.
- 9) Si un estudiante de género masculino y aprueba las materias de Lenguaje Ensamblador y Simulación es muy probable que también apruebe la asignatura de Análisis y Diseño de Sistemas.
- 10) Si un estudiante aprueba las materias de Lenguaje Ensamblador y Simulación es muy probable que también apruebe la asignatura de Análisis y Diseño de Sistemas.



- 11) Si un estudiante de género masculino y aprueba las materias de Análisis y Diseño de Sistemas es muy probable que también apruebe las asignaturas de Lenguaje Ensamblador y Simulación.
- 12) Si un estudiante es procedente al sector Urbano, no ejerce alguna actividad laboral y aprueba la materia Lenguaje Ensamblador es muy probable que también apruebe Análisis y Diseño de Sistemas y Simulación.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado con valores mayores a 1.16, además se ha filtrado con una confianza mayor al 85% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONF.	LIFT.	D	E	F
92. sector_procedencia=0 tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS-1 SIMULACION=1 107 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 93 conf:	0.87	lift:(1.18)>	lev:(0.04)	[14]	conv:(1.87)
99. sector_procedencia=0 tipo_colegio=2 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 108 ==> provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 93 conf:	0.86	lift:(1.18)>	lev:(0.04)	[13]	conv:(1.81)
120. sector_procedencia=0 tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 110 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 95 conf:	0.86	lift:(1.17)>	lev:(0.04)	[13]	conv:(1.8)
125. sector_procedencia=0 tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 110 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 93 conf:	0.85	lift:(1.17)>	lev:(0.04)	[13]	conv:(1.69)
136. sector_procedencia=0 trabaja=0 tipo_colegio=2 LENGUAJE_ENSAMBLADOR=1 102 ==> provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 87 conf:	0.85	lift:(1.16)>	lev:(0.04)	[12]	conv:(1.71)
138. tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 114 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 98 conf:	0.86	lift:(1.16)>	lev:(0.04)	[13]	conv:(1.76)
142. tipo_colegio=2 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 115 ==> provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 98 conf:	0.85	lift:(1.16)>	lev:(0.04)	[13]	conv:(1.71)
146. sector_procedencia=0 trabaja=0 tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 SIMULACION=1 99 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 85 conf:	0.86	lift:(1.16)>	lev:(0.04)	[11]	conv:(1.73)
161. sector_procedencia=0 trabaja=0 tipo_colegio=2 LENGUAJE_ENSAMBLADOR=1 SIMULACION=1 100 ==> provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 85 conf:	0.85	lift:(1.16)>	lev:(0.04)	[11]	conv:(1.67)
172. sector_procedencia=0 tipo_colegio=2 LENGUAJE_ENSAMBLADOR=1 112 ==> provincia_nacimiento=11 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 95 conf:	0.85	lift:(1.16)>	lev:(0.04)	[13]	conv:(1.67)
176. tipo_colegio=2 ANALISIS_Y_DISEÑO_DE_SISTEMAS=1 117 ==> provincia_nacimiento=11 LENGUAJE_ENSAMBLADOR=1 100 conf:	0.85	lift:(1.16)>	lev:(0.04)	[13]	conv:(1.7)

Figura 162. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios en un colegio Fiscomisional y aprueba Análisis y Diseño de Sistemas y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador.
- 2) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios en un colegio Fiscomisional y aprueba Lenguaje Ensamblador y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Análisis y Diseño de Sistemas.
- 3) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios en un colegio Fiscomisional y aprueba Análisis y Diseño de Sistemas, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador.
- 4) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios en un colegio Fiscomisional y aprueba Análisis y Diseño de Sistemas, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador y Simulación.





- 5) Si un estudiante es procedente del sector Urbano, no ejerce actividad laboral, ha culminado sus estudios en un colegio Fiscomisional y aprueba Lenguaje Ensamblador, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Análisis y Diseño de Sistemas.
- 6) Si un estudiante ha culminado sus estudios secundarios en una institución Fiscomisional y aprueba Análisis y Diseño de Sistemas y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador.
- 7) Si un estudiante ha culminado sus estudios secundarios en una institución Fiscomisional y aprueba Lenguaje Ensamblador y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Análisis y Diseño de Sistemas.
- 8) Si un estudiante es procedente del sector Urbano, no ejerce actividad laboral, ha culminado sus estudios en un colegio Fiscomisional y aprueba Análisis y Diseño de Sistemas y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador.
- 9) Si un estudiante es procedente del sector Urbano, no ejerce actividad laboral, ha culminado sus estudios en un colegio Fiscomisional y aprueba Lenguaje Ensamblador y Simulación, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Análisis y diseño de Sistemas.
- 10) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios en un colegio Fiscomisional y aprueba Lenguaje Ensamblador, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Análisis y Diseño de Sistemas.
- 11) Si un estudiante ha culminado sus estudios en un colegio Fiscomisional y aprueba Análisis y Diseño de Sistemas, es muy probable que pertenezca a la provincia de Loja y apruebe la materia de Lenguaje Ensamblador.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer únicamente a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes

dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es bajo, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo.

## Resultados obtenidos por Algoritmo A priori – RapidMiner

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, además, cuenta con la presencia de las variables que identifican si un estudiante trabaja o no y de qué provincia proviene o pertenece el estudiante. Además, la importancia del tipo de colegio del cual proviene el alumno, ya que, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 23 a 29 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

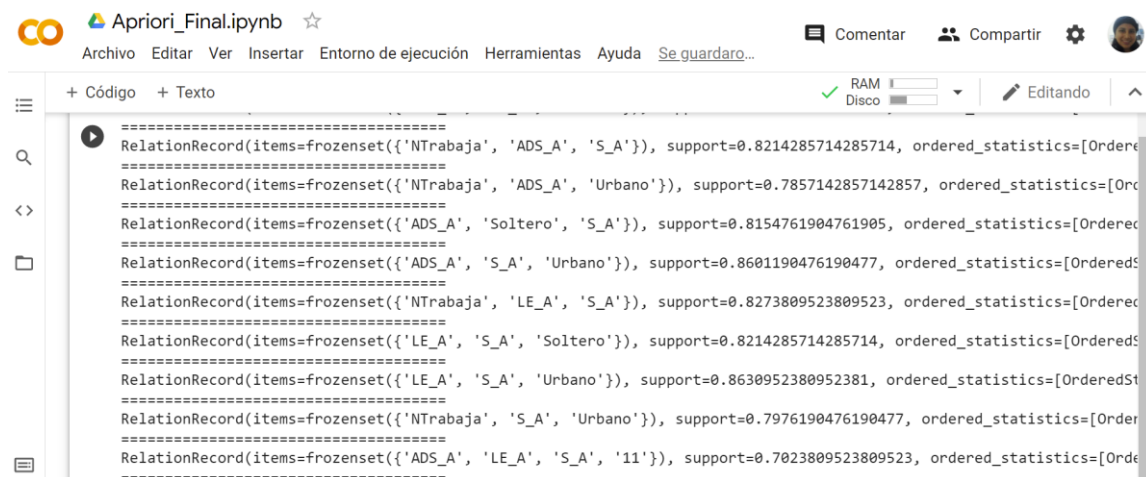
```
Association Rules

[trabaja] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.895)
[trabaja] --> [SIMULACION, LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.895)
[tipo_colegio] --> [ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.908)
[SIMULACION, LENGUAJE ENSAMBLADOR, sector_procedencia] --> [ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.909)
[genero] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.913)
[genero] --> [SIMULACION, LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.913)
[provincia_nacimiento] --> [SIMULACION] (confidence: 0.914)
[provincia_nacimiento] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.914)
[edad, tipo_colegio] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.914)
[edad, tipo_colegio] --> [SIMULACION, LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.914)
[SIMULACION, edad, tipo_colegio] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.914)
[SIMULACION, sector_procedencia] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.917)
[sector_procedencia] --> [SIMULACION] (confidence: 0.923)
[edad] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.927)
[edad] --> [SIMULACION, LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.927)
[SIMULACION, edad] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.927)
[tipo_colegio] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.929)
[SIMULACION, tipo_colegio] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.932)
[edad, provincia_nacimiento] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.933)
[edad, provincia_nacimiento] --> [SIMULACION, LENGUAJE ENSAMBLADOR] (confidence: 0.933)
[SIMULACION, edad, provincia_nacimiento] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.933)
[tipo_colegio, genero] --> [SIMULACION] (confidence: 0.935)
[tipo_colegio, genero] --> [LENGUAJE ENSAMBLADOR] (confidence: 0.935)
[tipo_colegio, genero] --> [SIMULACION, LENGUAJE ENSAMBLADOR] (confidence: 0.935)
[tipo_colegio, genero] --> [SIMULACION, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.935)
[edad, genero] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.935)
[tipo_colegio, genero] --> [LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.935)
[edad, genero] --> [SIMULACION, LENGUAJE ENSAMBLADOR, ANALISIS Y DISENO DE SISTEMAS] (confidence: 0.935)
```

Figura 163. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que sean aprobadas las materias el estudiante dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, si el estado civil del alumno es soltero también aprueban todas las materias, si los estudiantes pertenecen al sector Urbano y finalmente, la aprobación en paralelo con otras asignaturas, esto en conjunto con las características antes mencionadas.



```

=====
RelationRecord(items=frozenset({'NTrabaja', 'ADS_A', 'S_A'}), support=0.8214285714285714, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'NTrabaja', 'ADS_A', 'Urbano'}), support=0.7857142857142857, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'ADS_A', 'Soltero', 'S_A'}), support=0.8154761904761905, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'ADS_A', 'S_A', 'Urbano'}), support=0.8601190476190477, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'NTrabaja', 'LE_A', 'S_A'}), support=0.8273809523809523, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'LE_A', 'S_A', 'Soltero'}), support=0.8214285714285714, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'LE_A', 'S_A', 'Urbano'}), support=0.8630952380952381, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'NTrabaja', 'S_A', 'Urbano'}), support=0.7976190476190477, ordered_statistics=[Order
=====
RelationRecord(items=frozenset({'ADS_A', 'LE_A', 'S_A', '11'}), support=0.7023809523809523, ordered_statistics=[Order
=====

```

Figura 164. Reglas de asociación obtenidas por Python

## Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances	
0	117 ( 35%)
1	118 ( 35%)
2	8 ( 2%)
3	82 ( 24%)
4	11 ( 3%)

Figura 165. clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 1 y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Full Data (336.0)	Cluster#				
		0 (117.0)	1 (118.0)	2 (8.0)	3 (82.0)	4 (11.0)
edad	29	29	25	25	27	29
genero	0	1	0	0	0	0
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	1	0	0
tipo_colegio	1	1	1	2	2	2
ANALISIS_Y_DISEÑO_DE_SISTEMAS	1	1	1	0	1	1
LENGUAJE_ENSAMBLADOR	1	1	1	0	1	1
SIMULACION	1	1	1	0	1	1

Figura 166. Clúster obtenidos por Weka.

**Clúster\_0 (25%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al sexto ciclo, estas son: Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación.

**Clúster\_1 (35%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al sexto ciclo, estas son: Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación.

**Clúster\_2 (2%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que cuenten con una actividad laboral y que hayan finalizado su secundaria en una institución Fiscomisional, reprobarán todas las materias relacionadas a las

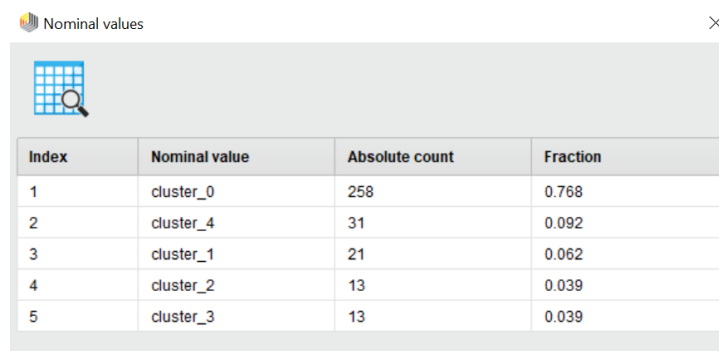
matemáticas correspondientes al sexto ciclo, estas son: Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación.

**Clúster\_3 (24%):** Estudiantes que tengan una edad de 27 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al sexto ciclo, estas son: Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación.

**Clúster\_4 (3%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al sexto ciclo, estas son: Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

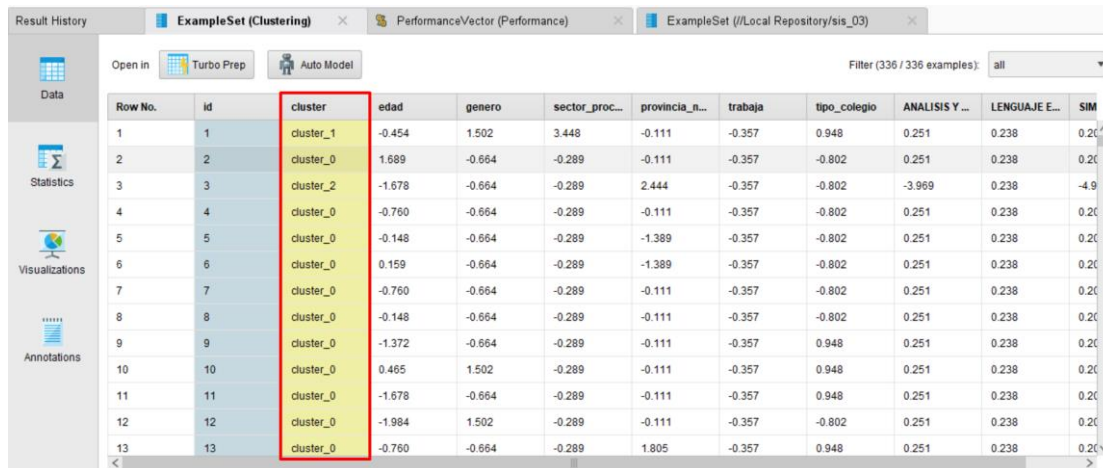
La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	258	0.768
2	cluster_4	31	0.092
3	cluster_1	21	0.062
4	cluster_2	13	0.039
5	cluster_3	13	0.039

Figura 167. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 336 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ANALISIS Y ...	LENGUAJE E...	SIM
1	1	cluster_1	-0.454	1.502	3.448	-0.111	-0.357	0.948	0.251	0.238	0.20
2	2	cluster_0	1.689	-0.664	-0.289	-0.111	-0.357	-0.802	0.251	0.238	0.20
3	3	cluster_2	-1.678	-0.664	-0.289	2.444	-0.357	-0.802	-3.969	0.238	-4.9
4	4	cluster_0	-0.760	-0.664	-0.289	-0.111	-0.357	-0.802	0.251	0.238	0.20
5	5	cluster_0	-0.148	-0.664	-0.289	-1.389	-0.357	-0.802	0.251	0.238	0.20
6	6	cluster_0	0.159	-0.664	-0.289	-1.389	-0.357	-0.802	0.251	0.238	0.20
7	7	cluster_0	-0.760	-0.664	-0.289	-0.111	-0.357	-0.802	0.251	0.238	0.20
8	8	cluster_0	-0.148	-0.664	-0.289	-0.111	-0.357	-0.802	0.251	0.238	0.20
9	9	cluster_0	-1.372	-0.664	-0.289	-0.111	-0.357	0.948	0.251	0.238	0.20
10	10	cluster_0	0.465	1.502	-0.289	-0.111	-0.357	0.948	0.251	0.238	0.20
11	11	cluster_0	-1.678	-0.664	-0.289	-0.111	-0.357	0.948	0.251	0.238	0.20
12	12	cluster_0	-1.984	1.502	-0.289	-0.111	-0.357	-0.802	0.251	0.238	0.20
13	13	cluster_0	-0.760	-0.664	-0.289	1.805	-0.357	0.948	0.251	0.238	0.20

Figura 168. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.

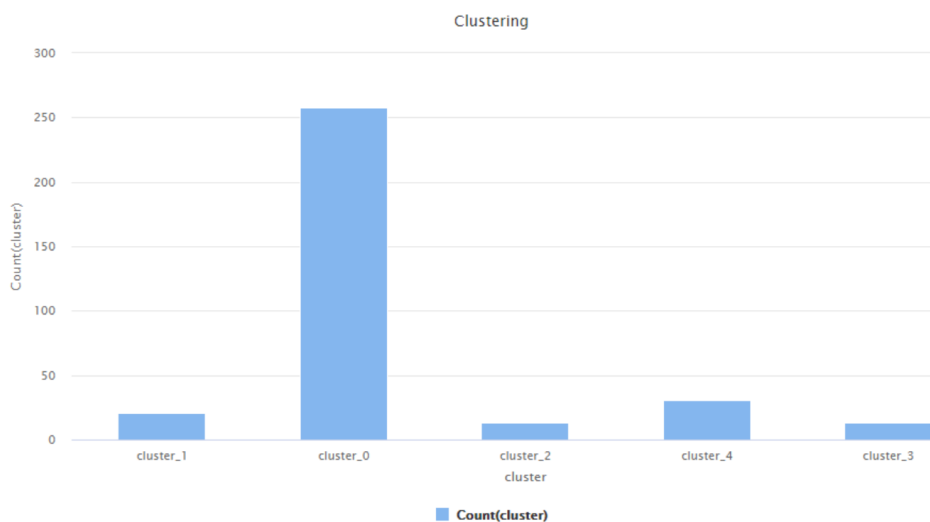


Figura 169. Representación gráfica de los clústers obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.



### k-Means - Centroid Table

Cluster	ANÁLISIS Y ...	edad	genero	LENGUAJE E...	provincia_n...	sector_proc...	SIMULACION	tipo_colegio	trabaja
Cluster 0	0.909	27.000	0.273	1	11.364	1	1	1.318	0.136
Cluster 1	0.992	27.705	0.318	1	10.398	-0	0.992	1.464	0.107
Cluster 2	1	26.300	0.500	0	11.600	0.100	0.600	1.300	0
Cluster 3	0.964	27.357	0.286	1	19.429	-0	1	1.321	0.107
Cluster 4	-0	25.333	0.067	0.467	12.600	0.200	0.533	1.933	0.267

Figura 170. Clúster obtenidos por Auto model de RapidMiner

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación, todas pertenecientes al sexto ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Rural, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación, todas pertenecientes al sexto ciclo, si tiene una edad de 28 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 2:** Estudiante aprueba algunas de las asignaturas relacionadas a las matemáticas como Análisis y Diseño de Sistemas y Simulación, reprueba la materia de Lenguaje Ensamblador, todas pertenecientes al sexto ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Zamora Chinchipe, perteneciente al sector Urbano, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 3:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Análisis y Diseño de Sistemas, Lenguaje ensamblador y Simulación, todas pertenecientes al sexto ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 4:** Estudiante aprueba únicamente la materia de Simulación y reprueba las asignaturas de Análisis y Diseño de Sistemas y Lenguaje ensamblador, todas pertenecientes al sexto ciclo, si tiene una edad de 25 años, es de género

masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y no ejerce alguna actividad laboral.

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide azul y verde, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color rojo hace referencia a estudiantes de género Femenino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, lo que indica que estudiantes de la Carrera de Ingeniería en Sistemas hay pocos estudiantes que culminan sus estudios en instituciones privadas.

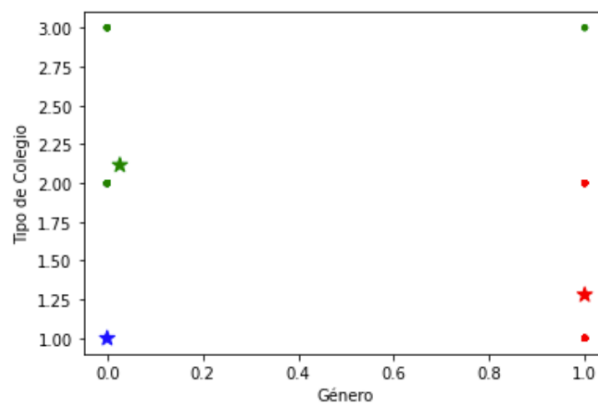


Figura 171. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Simulación** en donde los estudiantes aprueban a la asignatura si su edad es de 24, 28 y 33 años.



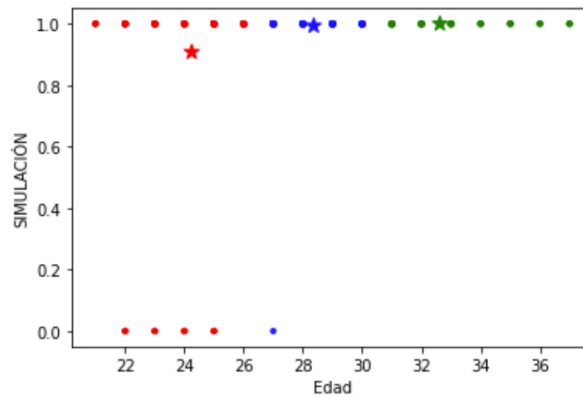


Figura 172. Relación Edad – Simulación

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Simulación**, donde se observa que un grupo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprueban la materia de Simulación.

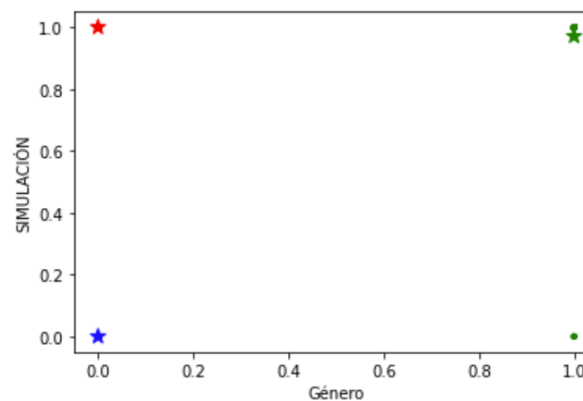


Figura 173. Relación Género – Simulación

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Lenguaje Ensamblador** y **Simulación**, en un clúster rojo se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde muy probable se apruebe a Simulación pero reprobará Lenguaje Ensamblador (clúster verde), y por último se muestra la reprobación de ambas asignaturas (clúster azul).

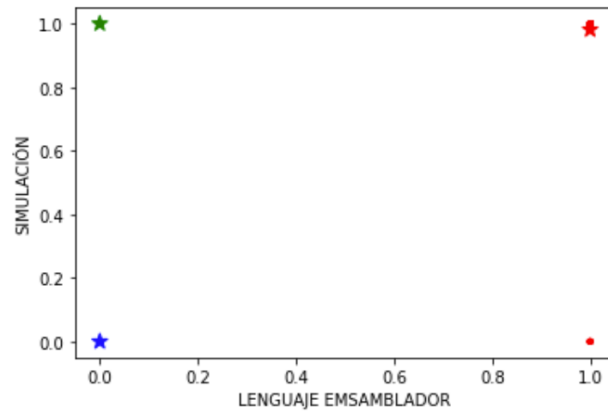


Figura 174. Relación Lenguaje Ensamblador - Simulación

## Anexo 10. Ingeniería en Sistemas – Séptimo ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **séptimo ciclo de la carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

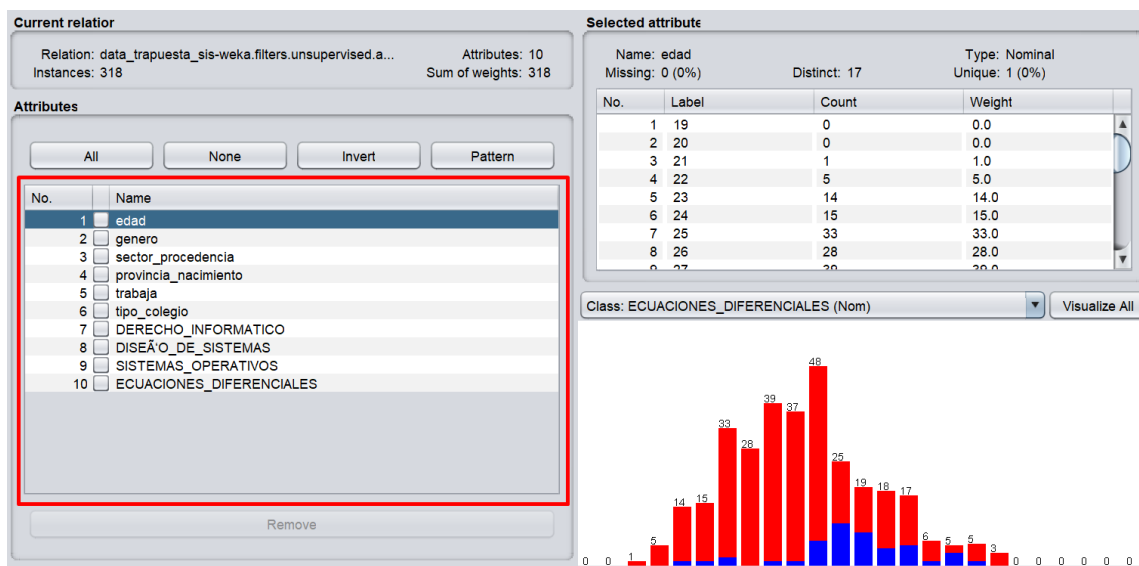
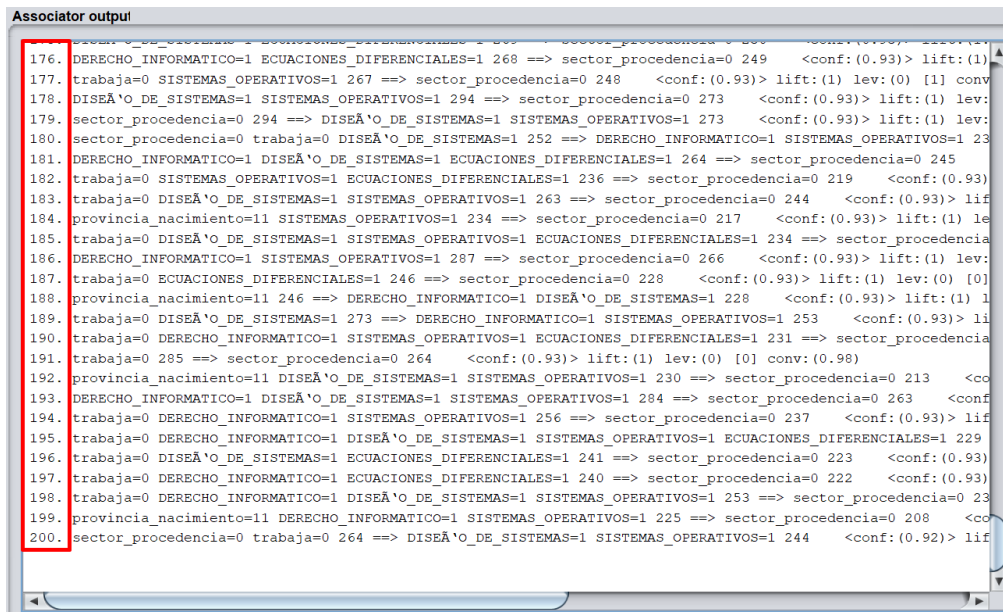


Figura 175. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a

tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.



```

176. DERECHO_INFORMATICO=1 ECUACIONES_DIFERENCIALES=1 268 ==> sector_procedencia=0 249 <conf:(0.93)> lift:(1)
177. trabaja=0 SISTEMAS_OPERATIVOS=1 267 ==> sector_procedencia=0 248 <conf:(0.93)> lift:(1) lev:(0) [1] conv
178. DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 294 ==> sector_procedencia=0 273 <conf:(0.93)> lift:(1) lev:
179. sector_procedencia=0 294 ==> DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 273 <conf:(0.93)> lift:(1) lev:
180. sector_procedencia=0 trabaja=0 DISEA'O_DE_SISTEMAS=1 252 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 23
181. DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 264 ==> sector_procedencia=0 245
182. trabaja=0 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 236 ==> sector_procedencia=0 219 <conf:(0.93)
183. trabaja=0 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 263 ==> sector_procedencia=0 244 <conf:(0.93)> lif
184. provincia_nacimiento=11 SISTEMAS_OPERATIVOS=1 234 ==> sector_procedencia=0 217 <conf:(0.93)> lift:(1) le
185. trabaja=0 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 234 ==> sector_procedencia
186. DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 287 ==> sector_procedencia=0 266 <conf:(0.93)> lift:(1) lev:
187. trabaja=0 ECUACIONES_DIFERENCIALES=1 246 ==> sector_procedencia=0 228 <conf:(0.93)> lift:(1) lev:(0) [0]
188. provincia_nacimiento=11 246 ==> DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 228 <conf:(0.93)> lift:(1) l
189. trabaja=0 DISEA'O_DE_SISTEMAS=1 273 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 253 <conf:(0.93)> li
190. trabaja=0 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 231 ==> sector_procedencia
191. trabaja=0 285 ==> sector_procedencia=0 264 <conf:(0.93)> lift:(1) lev:(0) [0] conv:(0.98)
192. provincia_nacimiento=11 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 230 ==> sector_procedencia=0 213 <co
193. DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 284 ==> sector_procedencia=0 263 <conf
194. trabaja=0 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 256 ==> sector_procedencia=0 237 <conf:(0.93)> lif
195. trabaja=0 DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 229
196. trabaja=0 DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 241 ==> sector_procedencia=0 223 <conf:(0.93)
197. trabaja=0 DERECHO_INFORMATICO=1 ECUACIONES_DIFERENCIALES=1 240 ==> sector_procedencia=0 222 <conf:(0.93)
198. trabaja=0 DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 253 ==> sector_procedencia=0 23
199. provincia_nacimiento=11 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 225 ==> sector_procedencia=0 208 <co
200. sector_procedencia=0 trabaja=0 264 ==> DISEA'O_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 244 <conf:(0.92)> lif
  
```

Figura 176. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza mayor a 95% y lift mayor a 1 respectivamente, con ello se obtiene un total de 17 reglas (Ver Figura 3).

REGLAS	CONF.	LIFT.	LEV.
15. sector_procedencia=0 provincia_nacimiento=11 DISEA'O_DE_SISTEMAS=1 216 ==> SISTEMAS_OPERATIVOS=1 213	conf:(0.99)>	lift:(1.05)	lev:(0.03)
59. SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 264 ==> DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 257	conf:(0.97)>	lift:(1.05)	lev:(0.04)
65. sector_procedencia=0 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 246 ==> DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 239	conf:(0.97)>	lift:(1.05)	lev:(0.03)
71. trabaja=0 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 236 ==> DERECHO_INFORMATICO=1 DISEA'O_DE_SISTEMAS=1 229	conf:(0.97)>	lift:(1.05)	lev:(0.03)
117. sector_procedencia=0 DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 250 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 239	conf:(0.96)>	lift:(1.06)	lev:(0.04)
121. DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 269 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 257	conf:(0.96)>	lift:(1.06)	lev:(0.04)
134. sector_procedencia=0 trabaja=0 DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 223 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 212	conf:(0.95)>	lift:(1.05)	lev:(0.03)
135. trabaja=0 DISEA'O_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 241 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 229	conf:(0.95)>	lift:(1.05)	lev:(0.04)
145. ECUACIONES_DIFERENCIALES=1 274 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 259	conf:(0.95)>	lift:(1.05)	lev:(0.04)
147. sector_procedencia=0 ECUACIONES_DIFERENCIALES=1 255 ==> DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 241	conf:(0.95)>	lift:(1.05)	lev:(0.03)

Figura 177. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

- 1) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y aprueba Diseño de Sistemas es muy probable que también apruebe la materia de Sistemas Operativos.
- 2) Si un estudiante aprueba materias como Sistemas Operativos y Ecuaciones Diferenciales probablemente también aprueben Derecho Informático y Diseño de Sistemas.
- 3) Si un estudiante es procedente del sector Urbano y aprueba Diseño de Sistemas y Ecuaciones Diferenciales es muy probable que también apruebe la materia de Derecho Informático y Diseño de Sistemas.
- 4) Si un estudiante no ejerce actividad laboral, aprueba Sistemas Operativos y Ecuaciones Diferenciales es muy probable que también apruebe la materia de Derecho Informático y Diseño de Sistemas.
- 5) Si un estudiante es procedente del sector Urbano, aprueba Diseño de Sistemas y Ecuaciones Diferenciales es muy probable que apruebe Derecho Informático y Sistemas Operativos.
- 6) Si un estudiante aprueba materias como Diseño de Sistemas y Ecuaciones Diferenciales probablemente también aprueben Derecho Informático y Sistemas Operativos.
- 7) Si un estudiante es procedente del sector Urbano, no ejerce actividad laboral y aprueba Diseño de Sistemas y Ecuaciones Diferenciales es muy probable que también apruebe la materia de Derecho Informático y Sistemas Operativos.
- 8) Si un estudiante no ejerce actividad laboral y aprueba Diseño de Sistemas y Ecuaciones Diferenciales es muy probable que también apruebe la materia de Derecho Informático y Sistemas Operativos.
- 9) Si un estudiante aprueba Ecuaciones Diferenciales es muy probable que apruebe Derecho Informático y Sistemas Operativos.
- 10) Si un estudiante es procedente del sector Urbano y aprueba Ecuaciones Diferenciales es muy probable que también apruebe Derecho Informático y Sistemas Operativos.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado con un valor de 1.20 en adelante y con una confiabilidad mayor al 85% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONF.	LIFT.	LEV.	
2. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 DISEÑO_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 82	0.87)	lift:(1.25)>	lev:(0.05)	[16]
4. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 83 conf:	0.88)	lift:(1.25)>	lev:(0.05)	[16]
6. tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 82	0.88)	lift:(1.25)>	lev:(0.05)	[16]
10. tipo_colegio=2 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 DISEÑO_DE_SISTEMAS=1 82	0.88)	lift:(1.23)>	lev:(0.05)	[15]
22. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 DISEÑO_DE_SISTEMAS=1 82 conf:	0.87)	lift:(1.22)>	lev:(0.05)	[14]
24. tipo_colegio=2 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 83 conf:	0.89)	lift:(1.21)>	lev:(0.05)	[14]
27. tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 ECUACIONES_DIFERENCIALES=1 92 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 82	0.89)	lift:(1.21)>	lev:(0.04)	[14]
38. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DISEÑO_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 82 conf:	0.87)	lift:(1.21)>	lev:(0.04)	[14]
40. tipo_colegio=2 DERECHO_INFORMATICO=1 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DISEÑO_DE_SISTEMAS=1 SISTEMAS_OPERATIVOS=1 82	0.87)	lift:(1.21)>	lev:(0.04)	[14]
51. tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 107 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 91 conf:	0.85)	lift:(1.2)>	lev:(0.05)	[15]
70. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 83 conf:	0.88)	lift:(1.2)>	lev:(0.04)	[13]
72. tipo_colegio=2 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 SISTEMAS_OPERATIVOS=1 83 conf:	0.88)	lift:(1.2)>	lev:(0.04)	[13]
74. tipo_colegio=2 DERECHO_INFORMATICO=1 ECUACIONES_DIFERENCIALES=1 94 ==> provincia_nacimiento=11 SISTEMAS_OPERATIVOS=1 83 conf:	0.88)	lift:(1.2)>	lev:(0.04)	[13]
85. sector_procedencia=0 tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 99 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 SISTEMAS_OPERATIVOS=1 84 conf:	0.85)	lift:(1.2)>	lev:(0.04)	[13]
90. tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 DERECHO_INFORMATICO=1 82 conf:	0.88)	lift:(1.2)>	lev:(0.04)	[13]
92. tipo_colegio=2 DISEÑO_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 SISTEMAS_OPERATIVOS=1 82 conf:	0.88)	lift:(1.2)>	lev:(0.04)	[13]
94. tipo_colegio=2 DERECHO_INFORMATICO=1 DISEÑO_DE_SISTEMAS=1 ECUACIONES_DIFERENCIALES=1 93 ==> provincia_nacimiento=11 SISTEMAS_OPERATIVOS=1 82	0.88)	lift:(1.2)>	lev:(0.04)	[13]

Figura 178. Reglas obtenidas en Weka con Lift mayor a 1

A continuación, se describe de manera general la particularidad que se presenta en cada uno de las premisas.

- 1) La mayoría de las premisas cuentan con la característica de tipo de colegio, el cual predomina las Instituciones Fiscomisionales.
- 2) Se necesita aprobar Ecuaciones Diferenciales para que se aprueben materias como Diseño de Sistemas, Sistemas operativos y Derecho informático y a esto se suma en su mayoría que los estudiantes sean oriundos de la provincia de Loja.

Las premisas que se han obtenido poseen una alta confiabilidad, así como también una alta predicción a futuro, y al ir de la mano, se pueden considerar para la toma de decisiones. Por otra parte cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es muy baja, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo.

### Resultados obtenidos por Algoritmo A priori – RapidMiner

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, sin embargo, en este caso se muestra una variable más predominante, que es el tipo de colegio del cual proviene el estudiante, si bien



es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 21 a 25 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### Association Rules

Association Rules

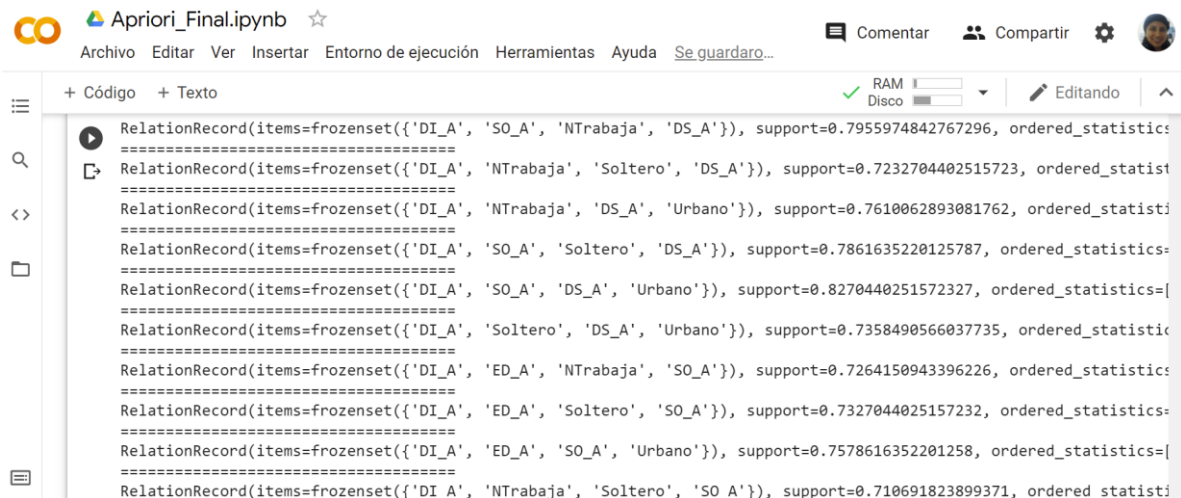
```
[DERECHO INFORMATICO, SISTEMAS OPERATIVOS, edad, tipo_colegio] --> [ECUACIONES DIFERENCIALES] (confidence: 0.800)
[DISENO DE SISTEMAS, edad, genero] --> [SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.804)
[genero] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.806)
[edad, tipo_colegio] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS] (confidence: 0.806)
[edad, genero] --> [ECUACIONES DIFERENCIALES] (confidence: 0.811)
[edad, genero] --> [DISENO DE SISTEMAS, ECUACIONES DIFERENCIALES] (confidence: 0.811)
[edad, genero] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS] (confidence: 0.811)
[edad, genero] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, SISTEMAS OPERATIVOS] (confidence: 0.811)
[tipo_colegio] --> [DISENO DE SISTEMAS, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.815)
[tipo_colegio] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.815)
[edad] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, SISTEMAS OPERATIVOS] (confidence: 0.815)
[genero] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.82)
[DISENO DE SISTEMAS, DERECHO INFORMATICO, edad, genero] --> [SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.822)
[tipo_colegio] --> [SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.823)
[tipo_colegio] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.823)
[genero] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, ECUACIONES DIFERENCIALES] (confidence: 0.827)
[edad] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS] (confidence: 0.829)
[DERECHO INFORMATICO, edad, genero] --> [ECUACIONES DIFERENCIALES] (confidence: 0.830)
[DERECHO INFORMATICO, edad, genero] --> [DISENO DE SISTEMAS, ECUACIONES DIFERENCIALES] (confidence: 0.830)
[tipo_colegio] --> [DISENO DE SISTEMAS, ECUACIONES DIFERENCIALES] (confidence: 0.831)
[tipo_colegio] --> [DISENO DE SISTEMAS, DERECHO INFORMATICO, ECUACIONES DIFERENCIALES] (confidence: 0.831)
[DISENO DE SISTEMAS, genero] --> [DERECHO INFORMATICO, SISTEMAS OPERATIVOS, ECUACIONES DIFERENCIALES] (confidence: 0.832)
[genero] --> [DERECHO INFORMATICO, ECUACIONES DIFERENCIALES] (confidence: 0.837)
[SISTEMAS OPERATIVOS, edad, genero] --> [ECUACIONES DIFERENCIALES] (confidence: 0.837)
[SISTEMAS OPERATIVOS, edad, genero] --> [DISENO DE SISTEMAS, ECUACIONES DIFERENCIALES] (confidence: 0.837)
[DISENO DE SISTEMAS, SISTEMAS OPERATIVOS, edad, genero] --> [ECUACIONES DIFERENCIALES] (confidence: 0.837)
[tipo_colegio] --> [ECUACIONES DIFERENCIALES] (confidence: 0.839)
[tipo_colegio] --> [DERECHO INFORMATICO, ECUACIONES DIFERENCIALES] (confidence: 0.839)
```

Figura 179. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3,4,5), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que estas sean aprobadas el estudiante debe dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, aprobará si el estado civil del alumno es soltero, si pertenece al sector Urbano y finalmente se da con la

aprobación en paralelo en conjunto con otras asignaturas, esto en relación a características mencionadas con anterioridad.



```

RelationRecord(items=frozenset({'DI_A', 'SO_A', 'NTrabaja', 'DS_A'}), support=0.7955974842767296, ordered_statistics=
=====
RelationRecord(items=frozenset({'DI_A', 'NTrabaja', 'Soltero', 'DS_A'}), support=0.7232704402515723, ordered_statist
=====
RelationRecord(items=frozenset({'DI_A', 'NTrabaja', 'DS_A', 'Urbano'}), support=0.7610062893081762, ordered_statisti
=====
RelationRecord(items=frozenset({'DI_A', 'SO_A', 'Soltero', 'DS_A'}), support=0.7861635220125787, ordered_statistics=
=====
RelationRecord(items=frozenset({'DI_A', 'SO_A', 'DS_A', 'Urbano'}), support=0.8270440251572327, ordered_statistics=[
=====
RelationRecord(items=frozenset({'DI_A', 'Soltero', 'DS_A', 'Urbano'}), support=0.7358490566037735, ordered_statistic
=====
RelationRecord(items=frozenset({'DI_A', 'ED_A', 'NTrabaja', 'SO_A'}), support=0.7264150943396226, ordered_statistics=
=====
RelationRecord(items=frozenset({'DI_A', 'ED_A', 'Soltero', 'SO_A'}), support=0.7327044025157232, ordered_statistics=
=====
RelationRecord(items=frozenset({'DI_A', 'ED_A', 'SO_A', 'Urbano'}), support=0.7578616352201258, ordered_statistics=[
=====
RelationRecord(items=frozenset({'DI_A', 'NTrabaja', 'Soltero', 'SO_A'}), support=0.710691823899371, ordered_statisti

```

Figura 180. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clúster obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances		
0	166	( 52%)
1	72	( 23%)
2	30	( 9%)
3	26	( 8%)
4	24	( 8%)

Figura 181. clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:



Final cluster centroids:

Attribute	Full Data (318.0)	Cluster#				
		0 (166.0)	1 (72.0)	2 (30.0)	3 (26.0)	4 (24.0)
edad	29	29	29	31	26	29
genero	0	0	1	0	0	1
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	7	11
trabaja	0	0	0	0	0	0
tipo_colegio	1	1	1	1	1	2
DERECHO_INFORMATICO	1	1	1	1	1	1
DISEÑO_DE_SISTEMAS	1	1	1	1	1	1
SISTEMAS_OPERATIVOS	1	1	1	1	1	1
ECUACIONES_DIFERENCIALES	1	1	1	0	1	1

Figura 182. Clúster obtenidos por Weka.

**Clúster\_0 (52%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al séptimo ciclo, estas son: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales

**Clúster\_1 (23%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al séptimo ciclo, estas son: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales

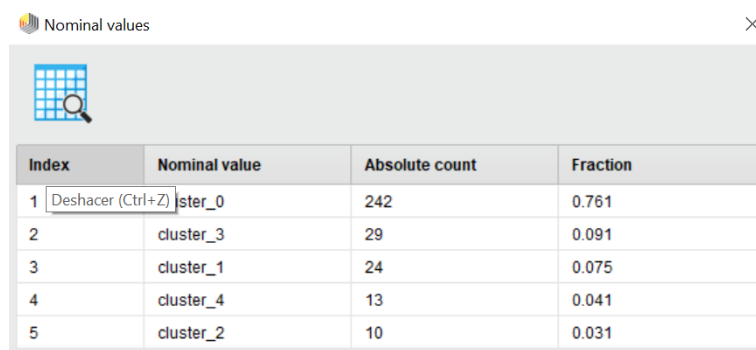
**Clúster\_2 (9%):** Estudiantes que tengan una edad de 31 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán tres materias relacionadas a las matemáticas, Derecho Informático, Diseño de Sistemas y Sistemas Operativos, mientras que reprobará Ecuaciones Diferenciales, todas correspondientes al séptimo ciclo.

**Clúster\_3 (8%):** Estudiantes que tengan una edad de 26 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de El Oro, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al séptimo ciclo, estas son: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales

**Clúster\_4 (8%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al séptimo ciclo, estas son: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	242	0.761
2	cluster_3	29	0.091
3	cluster_1	24	0.075
4	cluster_4	13	0.041
5	cluster_2	10	0.031

Figura 183. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 318 instancias están clasificadas o agrupadas a su clúster más cercano.

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	DERECHO IN...	DISEÑO DE ...	SIS
1	1	cluster_1	-0.707	1.496	3.494	-0.081	-0.340	1.010	0.222	0.206	0.22
2	2	cluster_4	1.463	-0.666	-0.285	-0.081	-0.340	-0.755	-4.487	-4.836	-3.8
3	3	cluster_0	-1.017	-0.666	-0.285	-0.081	-0.340	-0.755	0.222	0.206	0.25
4	4	cluster_0	-0.397	-0.666	-0.285	-1.319	-0.340	-0.755	0.222	0.206	0.25
5	5	cluster_0	-0.087	-0.666	-0.285	-1.319	-0.340	-0.755	0.222	0.206	0.25
6	6	cluster_0	-1.017	-0.666	-0.285	-0.081	-0.340	-0.755	0.222	0.206	0.25
7	7	cluster_0	-0.397	-0.666	-0.285	-0.081	-0.340	-0.755	0.222	0.206	0.25
8	8	cluster_0	0.223	1.496	-0.285	-0.081	-0.340	1.010	0.222	0.206	0.25
9	9	cluster_0	-1.947	-0.666	-0.285	-0.081	-0.340	1.010	0.222	0.206	0.25
10	10	cluster_0	-1.017	-0.666	-0.285	1.777	-0.340	1.010	0.222	0.206	0.25
11	11	cluster_0	-0.087	-0.666	-0.285	-1.319	-0.340	1.010	0.222	0.206	0.25
12	12	cluster_2	0.533	-0.666	-0.285	-0.081	-0.340	1.010	-4.487	0.206	0.25
13	13	cluster_0	0.843	1.496	-0.285	-0.081	-0.340	1.010	0.222	0.206	0.25

Figura 184. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.

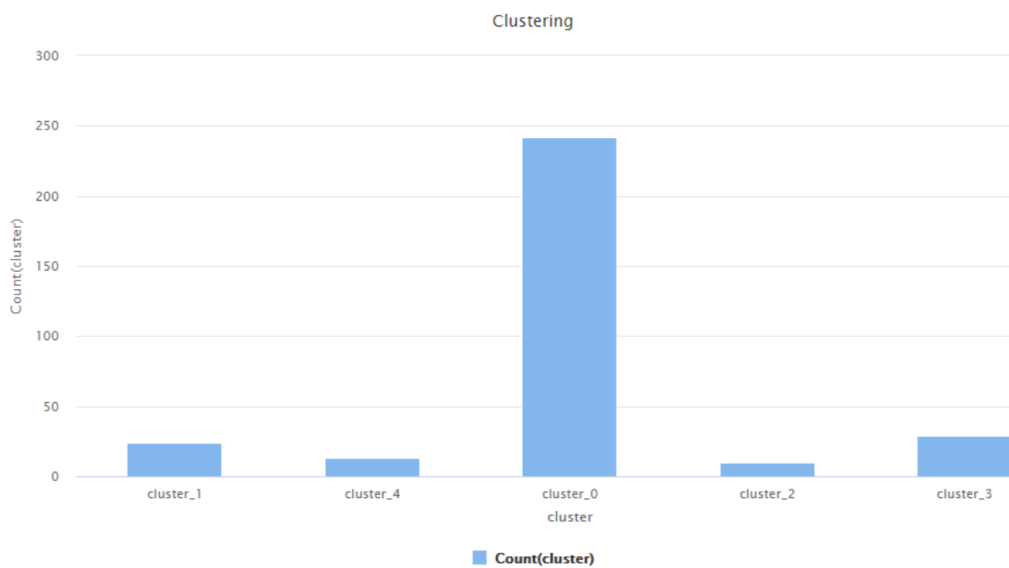


Figura 185. Representación gráfica de los clústers obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo queremos ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

### k-Means - Centroid Table

Cluster	DERECHO L...	DISENO DE...	ECUACION...	edad	genero	provincia_...	sector_pr...	SISTEMAS...	tipo_colegio	trabaja
Cluster 0	1	1	0.987	27.348	0	10.903	0.097	1	1.516	0
Cluster 1	0.615	0	0.385	31.923	0.231	11.462	0	0.308	1.538	0.077
Cluster 2	1	1	0.903	28.065	0.097	11.581	0.065	1	1.290	1.000
Cluster 3	1	1	0.636	29.000	0.182	10.909	0.273	0	1.364	0.091
Cluster 4	0.907	1	0.750	29.167	0.833	11.694	0.037	1	1.333	0

Figura 186. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales, todas pertenecientes al séptimo ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, ejerce actividad laboral.

**Clúster 1:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales, todas pertenecientes al séptimo ciclo, si tiene una edad de 31 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, ejerce actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales, todas pertenecientes al séptimo ciclo, si tiene una edad de 28 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal y ejerce actividad laboral extra a sus actividades académicas.

**Clúster 3:** Estudiante aprueba ciertas asignaturas relacionadas a las matemáticas como: Derecho Informático, Diseño de Sistemas y reprueba Sistemas Operativos y Ecuaciones Diferenciales, todas pertenecientes al séptimo ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

Fiscal, y únicamente se dedica a sus actividades académicas, es decir, ejerce actividad laboral.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como: Derecho Informático, Diseño de Sistemas, Sistemas Operativos y Ecuaciones Diferenciales, todas pertenecientes al séptimo ciclo, si tiene una edad de 29 años, es de género femenino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, ejerce actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide verde y rojo, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color azul hace referencia a estudiantes de género Femenino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, lo que indica que estudiantes de la Carrera de Ingeniería en Sistemas hay pocos estudiantes que culminan sus estudios en instituciones privadas.

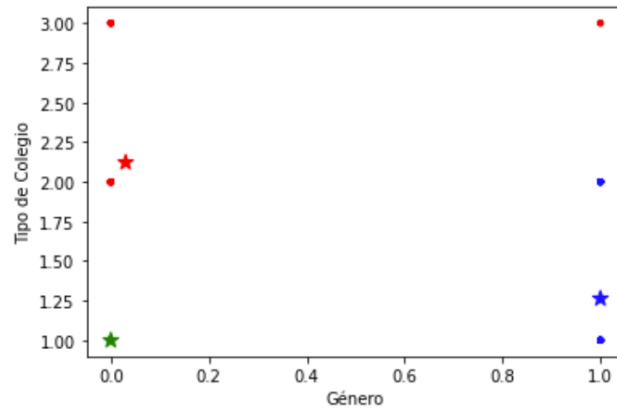


Figura 187. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Ecuaciones Diferenciales** en donde los estudiantes aprueban a la asignatura si su edad es de 25, 28 y 33 años.

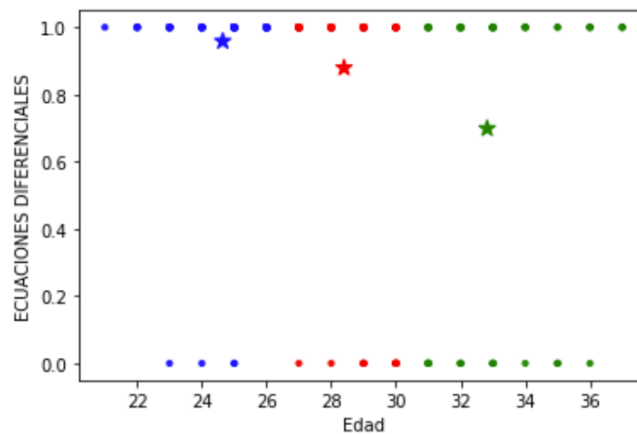


Figura 188. Relación Edad – Ecuaciones Diferenciales

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Ecuaciones Diferenciales**, donde se observa que un grupo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprueban Ecuaciones Diferenciales.

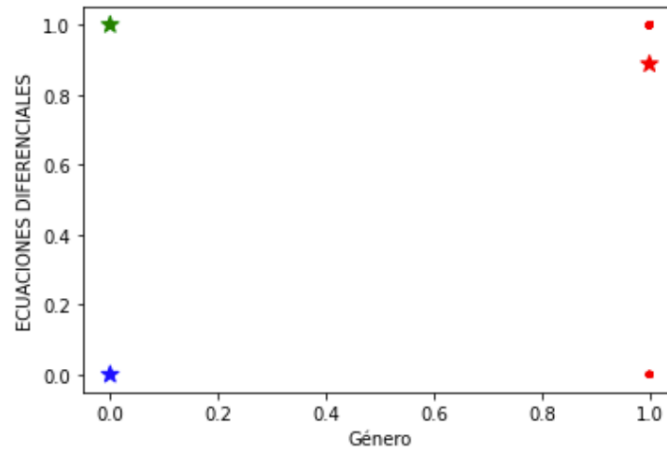


Figura 189. Relación Género – Ecuaciones Diferenciales

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Sistemas Operativos** y **Ecuaciones Diferenciales**, en el clúster verde se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde apruebe Sistemas Operativos pero reprueba Ecuaciones Diferenciales (clúster azul), y por último se muestra la reprobación de Sistemas Operativos y con una probabilidad del 50 % que apruebe Ecuaciones Diferenciales (clúster rojo).

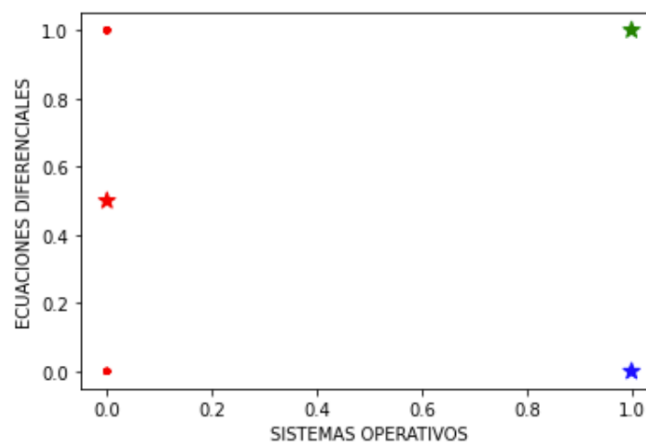


Figura 190. Relación Sistemas Operativos - Ecuaciones Diferenciales



## Anexo 11. Ingeniería en Sistemas – Octavo ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **octavo ciclo de la carrera de Ingeniería en Sistemas**, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clústers, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 100 reglas.

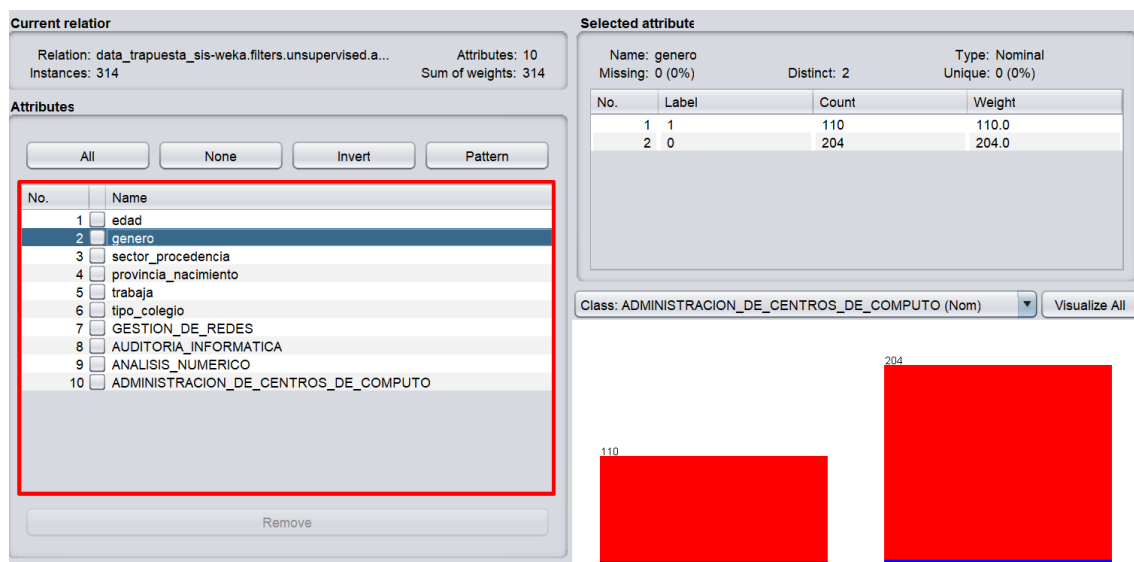
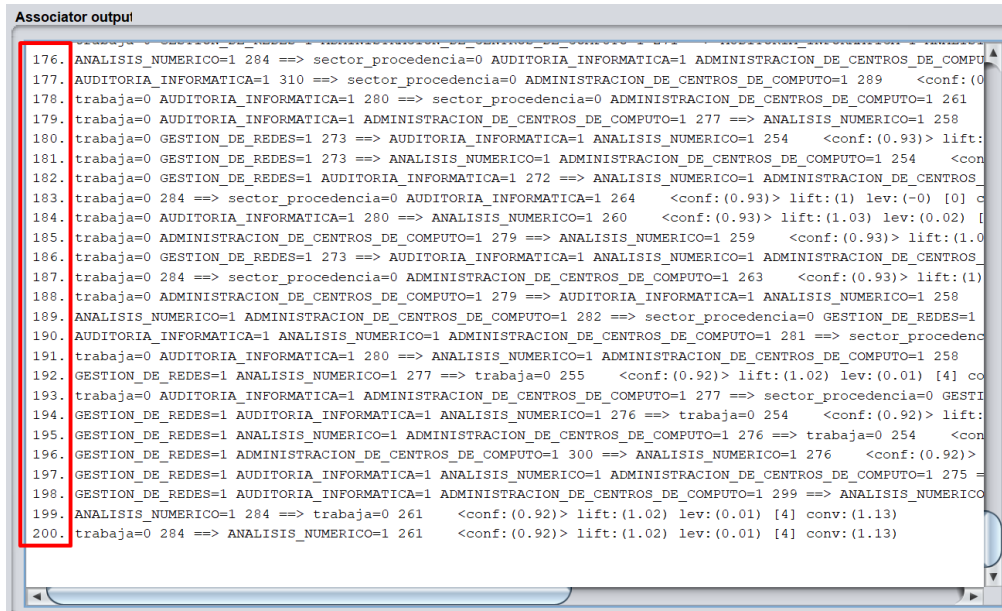


Figura 191. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a



tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.



```
Associator output
176. ANALISIS_NUMERICO=1 284 ==> sector_procedencia=0 AUDITORIA_INFORMATICA=1 ADMINISTRACION_DE_CENTROS_DE_COMPU
177. AUDITORIA_INFORMATICA=1 310 ==> sector_procedencia=0 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 289 <conf: (0
178. trabaja=0 AUDITORIA_INFORMATICA=1 280 ==> sector_procedencia=0 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 261
179. trabaja=0 AUDITORIA_INFORMATICA=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 277 ==> ANALISIS_NUMERICO=1 258
180. trabaja=0 GESTION_DE_REDES=1 273 ==> AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 254 <conf: (0.93)> lift:
181. trabaja=0 GESTION_DE_REDES=1 273 ==> ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 254 <con
182. trabaja=0 GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 272 ==> ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS
183. trabaja=0 284 ==> sector_procedencia=0 AUDITORIA_INFORMATICA=1 264 <conf: (0.93)> lift: (1) lev: (-0) [0] c
184. trabaja=0 AUDITORIA_INFORMATICA=1 280 ==> ANALISIS_NUMERICO=1 260 <conf: (0.93)> lift: (1.03) lev: (0.02) [
185. trabaja=0 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 279 ==> ANALISIS_NUMERICO=1 259 <conf: (0.93)> lift: (1.0
186. trabaja=0 GESTION_DE_REDES=1 273 ==> AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS
187. trabaja=0 284 ==> sector_procedencia=0 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 263 <conf: (0.93)> lift: (1)
188. trabaja=0 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 279 ==> AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 258
189. ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 282 ==> sector_procedencia=0 GESTION_DE_REDES=1
190. AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 281 ==> sector_procedenc
191. trabaja=0 AUDITORIA_INFORMATICA=1 280 ==> ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 258
192. GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 277 ==> trabaja=0 255 <conf: (0.92)> lift: (1.02) lev: (0.01) [4] co
193. trabaja=0 AUDITORIA_INFORMATICA=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 277 ==> sector_procedencia=0 GESTI
194. GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 276 ==> trabaja=0 254 <conf: (0.92)> lift:
195. GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 276 ==> trabaja=0 254 <con
196. GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 300 ==> ANALISIS_NUMERICO=1 276 <conf: (0.92)>
197. GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 275 =
198. GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 299 ==> ANALISIS_NUMERICO
199. ANALISIS_NUMERICO=1 284 ==> trabaja=0 261 <conf: (0.92)> lift: (1.02) lev: (0.01) [4] conv: (1.13)
200. trabaja=0 284 ==> ANALISIS_NUMERICO=1 261 <conf: (0.92)> lift: (1.02) lev: (0.01) [4] conv: (1.13)
```

Figura 192. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza igual a 100% y lift mayor a 1 respectivamente, se obtiene un total de 16 reglas (Ver Figura 3).

REGLAS	CONF.	LIFT.		
1. GESTION_DE_REDES=1 302 ==> AUDITORIA_INFORMATICA=1 301	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
2. GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 300 ==> AUDITORIA_INFORMATICA=1 299	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
3. sector_procedencia=0 GESTION_DE_REDES=1 285 ==> AUDITORIA_INFORMATICA=1 284	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
4. ANALISIS_NUMERICO=1 284 ==> AUDITORIA_INFORMATICA=1 283	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
5. sector_procedencia=0 GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 283 ==> AUDITORIA_INFORMATICA=1 282	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
6. ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 282 ==> AUDITORIA_INFORMATICA=1 281	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
7. GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 277 ==> AUDITORIA_INFORMATICA=1 276	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
8. GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 277 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 276	conf:(1)>	lift:(1.01)	lev:(0.01)	[3]
9. GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 276 ==> AUDITORIA_INFORMATICA=1 275	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
10. GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 276 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 275	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
11. trabaja=0 GESTION_DE_REDES=1 273 ==> AUDITORIA_INFORMATICA=1 272	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
12. trabaja=0 GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 271 ==> AUDITORIA_INFORMATICA=1 270	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
13. sector_procedencia=0 ANALISIS_NUMERICO=1 268 ==> AUDITORIA_INFORMATICA=1 267	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
14. sector_procedencia=0 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 266 ==> AUDITORIA_INFORMATICA=1 265	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
15. trabaja=0 ANALISIS_NUMERICO=1 261 ==> AUDITORIA_INFORMATICA=1 260	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
16. sector_procedencia=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 261 ==> AUDITORIA_INFORMATICA=1 260	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
17. sector_procedencia=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 261 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 260	conf:(1)>	lift:(1.01)	lev:(0.01)	[3]
18. sector_procedencia=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 260 ==> AUDITORIA_INFORMATICA=1 259	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
19. sector_procedencia=0 GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 260 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 259	conf:(1)>	lift:(1.01)	lev:(0.01)	[3]
20. trabaja=0 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 259 ==> AUDITORIA_INFORMATICA=1 258	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
21. sector_procedencia=0 trabaja=0 GESTION_DE_REDES=1 258 ==> AUDITORIA_INFORMATICA=1 257	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
22. sector_procedencia=0 trabaja=0 GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 256 ==> AUDITORIA_INFORMATICA=1 255	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
23. trabaja=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 255 ==> AUDITORIA_INFORMATICA=1 254	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
24. trabaja=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 255 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 254	conf:(1)>	lift:(1.01)	lev:(0.01)	[2]
25. trabaja=0 GESTION_DE_REDES=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 254 ==> AUDITORIA_INFORMATICA=1 253	conf:(1)>	lift:(1.01)	lev:(0.01)	[3]
26. trabaja=0 GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 254 ==> ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 253	conf:(1)>	lift:(1.01)	lev:(0.01)	[3]

Figura 193. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante aprueba Gestión de Redes muy probablemente también aprobará Auditoria Informática.
- 2) Si un estudiante aprueba Gestión de Redes y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 3) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes muy probablemente también aprobará Auditoria Informática.
- 4) Si un estudiante aprueba Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 5) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes y Administración de Centros de Cómputo, muy probablemente también aprobará Auditoria Informática.
- 6) Si un estudiante aprueba Análisis Numérico y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 7) Si un estudiante aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 8) Si un estudiante aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Administración de Centros de Cómputo.
- 9) Si un estudiante aprueba Gestión de Redes, Análisis Numérico y Administración de Centros de Cómputo, muy probablemente también aprobará Auditoria Informática.



- 10) Si un estudiante es procedente del sector Urbano, aprueba Gestión de Redes, Análisis Numérico y Administración de Centros de Cómputo, muy probablemente también aprobará Auditoria Informática.
- 11) Si un estudiante no ejerce actividad laboral y aprueba Gestión de Redes es muy probable que también apruebe Auditoria Informática.
- 12) Si un estudiante no ejerce actividad laboral y aprueba Gestión de Redes y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 13) Si un estudiante es procedente del sector Urbano y aprueba Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 14) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 15) Si un estudiante no ejerce actividad laboral y aprueba Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 16) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 17) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Administración de centros de Cómputo.
- 18) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes, Análisis Numérico y Administración de centros de Cómputo muy probablemente también aprobará Auditoria Informática.
- 19) Si un estudiante es procedente del sector Urbano y aprueba Gestión de Redes, Auditoria Informática y Análisis Numérico, muy probablemente también aprobará Administración de centros de Cómputo.
- 20) Si un estudiante ejerce actividad laboral y aprueba Gestión de Redes y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 21) Si un estudiante es procedente del sector Urbano, ejerce actividad laboral y aprueba Gestión de Redes muy probablemente también aprobará Auditoria Informática.

- 22) Si un estudiante es procedente del sector Urbano, ejerce actividad laboral y aprueba Gestión de Redes y Administración de Centros de Computo muy probablemente también aprobará Auditoria Informática.
- 23) Si un estudiante ejerce actividad laboral y aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Auditoria Informática.
- 24) Si un estudiante ejerce actividad laboral y aprueba Gestión de Redes y Análisis Numérico muy probablemente también aprobará Administración de Centros de Cómputo.
- 25) Si un estudiante ejerce actividad laboral y aprueba Gestión de Redes, Análisis Numérico y Administración de Centros de Cómputo, muy probablemente también aprobará Auditoria Informática.
- 26) Si un estudiante ejerce actividad laboral y aprueba Gestión de Redes, Análisis Numérico y Auditoria Informática, muy probablemente también aprobará Administración de Centros de Cómputo.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado obteniendo premisas con valores mayor a 1.20 y una confiabilidad del 80 % (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONFI.	LIFT.	
2. genero=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 98 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 80 conf:	0.82	lift:(1.28)>	lev:(0.06) [17]
4. genero=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 97 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 79	0.81	lift:(1.28)>	lev:(0.05) [17]
6. genero=1 ANALISIS_NUMERICO=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 98 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 79	0.81	lift:(1.27)>	lev:(0.05) [16]
8. genero=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 108 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 87 conf:	0.81	lift:(1.26)>	lev:(0.06) [18]
9. genero=1 ANALISIS_NUMERICO=1 100 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 80 conf:	0.8	lift:(1.26)>	lev:(0.05) [16]
12. genero=1 AUDITORIA_INFORMATICA=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 107 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 86 conf:	0.8	lift:(1.26)>	lev:(0.06) [17]
13. genero=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 99 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 79	0.8	lift:(1.26)>	lev:(0.05) [16]
16. genero=1 ADMINISTRACION_DE_CENTROS_DE_COMPUTO=1 108 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 AUDITORIA_INFORMATICA=1 86 conf:	0.8	lift:(1.26)>	lev:(0.06) [17]
18. genero=1 ANALISIS_NUMERICO=1 100 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 80 conf:	0.8	lift:(1.26)>	lev:(0.05) [16]
20. genero=1 AUDITORIA_INFORMATICA=1 ANALISIS_NUMERICO=1 99 ==> sector_procedencia=0 provincia_nacimiento=11 trabaja=0 GESTION_DE_REDES=1 79 conf:	0.8	lift:(1.25)>	lev:(0.05) [15]

Figura 194. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante de género femenino y aprueba las asignaturas de Análisis Numérico y Administración de centros de Cómputo es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.
- 2) Si un estudiante de género femenino y aprueba las asignaturas de Auditoria Informática, Análisis Numérico y Administración de centros de Cómputo, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.



- 3) Si un estudiante de género femenino y aprueba las asignaturas de Análisis Numérico y Administración de centros de Cómputo, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes y Auditoria Informática.
- 4) Si un estudiante de género femenino y aprueba Administración de centros de Cómputo, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.
- 5) Si un estudiante de género femenino y aprueba Análisis Numérico, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes y Auditoria Informática.
- 6) Si un estudiante de género femenino y aprueba las asignaturas de Auditoria Informática y Administración de centros de Cómputo, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.
- 7) Si un estudiante de género femenino y aprueba las asignaturas de Auditoria Informática y Análisis Numérico, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes y Administración de Centros de Cómputo.
- 8) Si un estudiante de género femenino y aprueba Administración de centros de Cómputo, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes y Auditoria Informática.
- 9) Si un estudiante de género femenino y aprueba Análisis Numérico, es muy probable que sea procedente del sector Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.
- 10) Si un estudiante de género femenino y aprueba Auditoria Informática y Análisis Numérico, es muy probable que sea procedente del sector





Urbano, pertenezca a la provincia de Loja, no cuente con actividad laboral y apruebe la materia de Gestión de Redes.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular dentro de este ciclo es muy alta, esto se deduce a que si bien es cierto son asignaturas relacionadas a las matemáticas no existe la presión de la “lógica” como en el caso de la programación, entre otras.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, además se sigue reflejando el tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 31 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la ausencia de la característica género. Esto se ve reflejado en la Figura 5.

## AssociationRules

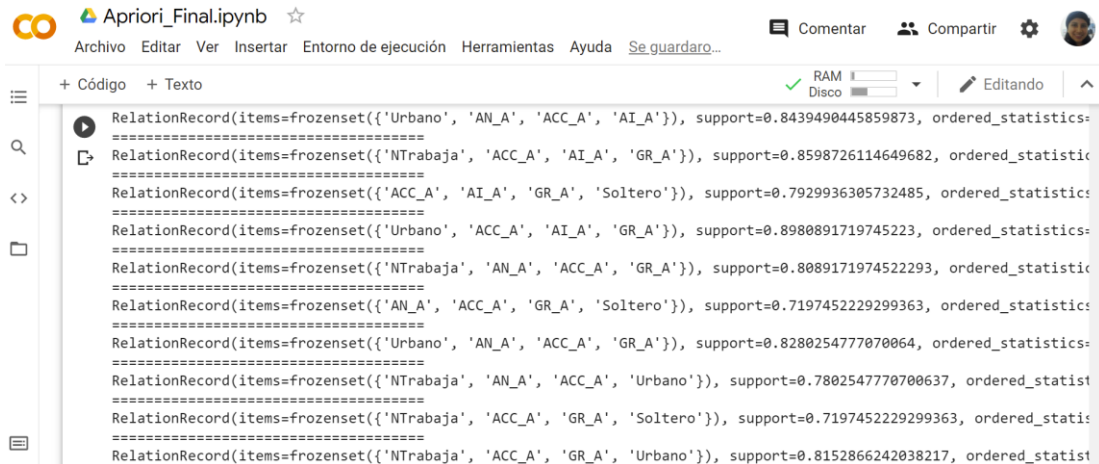
### Association Rules

```
[tipo_colegio, edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.816)
[tipo_colegio, edad] --> [AUDITORIA INFORMATICA, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.816)
[tipo_colegio, edad] --> [ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.816)
[tipo_colegio, edad] --> [AUDITORIA INFORMATICA, ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.816)
[ADMINISTRACION DE CENTROS DE COMPUTO, tipo_colegio, edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.833)
[ADMINISTRACION DE CENTROS DE COMPUTO, tipo_colegio, edad] --> [AUDITORIA INFORMATICA, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.833)
[AUDITORIA INFORMATICA, tipo_colegio, edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.851)
[AUDITORIA INFORMATICA, tipo_colegio, edad] --> [ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.851)
[AUDITORIA INFORMATICA, ADMINISTRACION DE CENTROS DE COMPUTO, tipo_colegio, edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.851)
[edad] --> [ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.855)
[edad] --> [AUDITORIA INFORMATICA, ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.855)
[tipo_colegio, edad] --> [ANALISIS NUMERICO] (confidence: 0.857)
[tipo_colegio, edad] --> [AUDITORIA INFORMATICA, ANALISIS NUMERICO] (confidence: 0.857)
[tipo_colegio, edad] --> [ADMINISTRACION DE CENTROS DE COMPUTO, ANALISIS NUMERICO] (confidence: 0.857)
[tipo_colegio, edad] --> [AUDITORIA INFORMATICA, ADMINISTRACION DE CENTROS DE COMPUTO, ANALISIS NUMERICO] (confidence: 0.857)
[edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.864)
[edad] --> [AUDITORIA INFORMATICA, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.864)
[tipo_colegio] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.87)
[tipo_colegio] --> [AUDITORIA INFORMATICA, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.870)
[tipo_colegio] --> [ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.870)
[tipo_colegio] --> [AUDITORIA INFORMATICA, ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.870)
[AUDITORIA INFORMATICA, edad] --> [ADMINISTRACION DE CENTROS DE COMPUTO, GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.870)
[ADMINISTRACION DE CENTROS DE COMPUTO, tipo_colegio, edad] --> [ANALISIS NUMERICO] (confidence: 0.875)
[ADMINISTRACION DE CENTROS DE COMPUTO, tipo_colegio, edad] --> [AUDITORIA INFORMATICA, ANALISIS NUMERICO] (confidence: 0.875)
[AUDITORIA INFORMATICA, edad] --> [GESTION DE REDES, ANALISIS NUMERICO] (confidence: 0.880)
```

Figura 195. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que estas sean aprobadas el estudiante debe dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, aprobará a las mismas si el estado civil del alumno es soltero, si pertenece al sector Urbano y finalmente se da con la aprobación en paralelo en conjunto con otras asignaturas, esto en relación a características mencionadas con anterioridad, además de ver la relación entre las materias de Gestión de Redes con Administración de Centros de Cómputo, las mismas que guardan estrecha relación.



```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardaro...
+ Código + Texto
RelationRecord(items=frozenset({'Urbano', 'AN_A', 'ACC_A', 'AI_A'}), support=0.8439490445859873, ordered_statistics=
RelationRecord(items=frozenset({'NTrabaja', 'ACC_A', 'AI_A', 'GR_A'}), support=0.8598726114649682, ordered_statistic
RelationRecord(items=frozenset({'ACC_A', 'AI_A', 'GR_A', 'Soltero'}), support=0.7929936305732485, ordered_statistics
RelationRecord(items=frozenset({'Urbano', 'ACC_A', 'AI_A', 'GR_A'}), support=0.8980891719745223, ordered_statistics
RelationRecord(items=frozenset({'NTrabaja', 'AN_A', 'ACC_A', 'GR_A'}), support=0.8089171974522293, ordered_statistic
RelationRecord(items=frozenset({'AN_A', 'ACC_A', 'GR_A', 'Soltero'}), support=0.7197452229299363, ordered_statistics
RelationRecord(items=frozenset({'Urbano', 'AN_A', 'ACC_A', 'GR_A'}), support=0.8280254777070064, ordered_statistics
RelationRecord(items=frozenset({'NTrabaja', 'AN_A', 'ACC_A', 'Urbano'}), support=0.7802547770700637, ordered_statist
RelationRecord(items=frozenset({'NTrabaja', 'ACC_A', 'GR_A', 'Soltero'}), support=0.7197452229299363, ordered_statist
RelationRecord(items=frozenset({'NTrabaja', 'ACC_A', 'GR_A', 'Urbano'}), support=0.8152866242038217, ordered_statist

```

Figura 196. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústers obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clústers, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances

Cluster	Instances	Percentage
0	107	( 34%)
1	163	( 52%)
2	3	( 1%)
3	39	( 12%)
4	2	( 1%)

Figura 197. clústeres con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 3. La Figura 7, indica las características que componen a cada uno de ellos, seguidamente se describe a cada clúster:

Final cluster centroids:

Attribute	Full Data (314.0)	Cluster#				
		0 (107.0)	1 (163.0)	2 (3.0)	3 (39.0)	4 (2.0)
edad	29	27	25	23	29	28
genero	0	1	0	0	1	1
sector_procedencia	0	0	0	0	0	1
provincia_nacimiento	11	11	11	1	11	19
trabaja	0	0	0	0	0	0
tipo_colegio	1	1	1	1	2	1
GESTION_DE_REDES	1	1	1	1	1	1
AUDITORIA_INFORMATICA	1	1	1	1	1	1
ANALISIS_NUMERICO	1	1	1	1	1	1
ADMINISTRACION_DE_CENTROS_DE_COMPUTO	1	1	1	1	1	1

Figura 198. Clúster obtenidos por Weka.





UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

**Clúster\_0 (34%):** Estudiantes que tengan una edad de 27 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al octavo ciclo, estas son: Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo.

**Clúster\_1 (52%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al octavo ciclo, estas son: Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo.

**Clúster\_2 (1%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Azuay, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al octavo ciclo, estas son: Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo.

**Clúster\_3 (12%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al octavo ciclo, estas son: Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo.

**Clúster\_4 (1%):** Estudiantes que tengan una edad de 28 años, pertenezcan al género femenino, sean del sector Rural y provenientes de la provincia de Zamora Chinchipe, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en

una institución Fiscal, reprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Gestión de Redes, Auditoria Informática, Análisis Numérico y Administración de Centros de Cómputo.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total del 5 clústers.

Nominal values ×

Index	Nominal value	Absolute count	Fraction
1	cluster_0	237	0.755
2	cluster_4	40	0.127
3	cluster_3	30	0.096
4	cluster_2	4	0.013
5	cluster_1	3	0.010

Figura 199. Clústers con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 314 instancias están clasificadas o agrupadas a su clúster más cercano.

Result History | ExampleSet (Clustering) | PerformanceVector (Performance) | ExampleSet (//Local Repository/sis\_03)

Open in Turbo Prep | Auto Model | Filter (314 / 314 examples): all

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	GESTION DE ...	AUDITORIA L...	ANM
1	1	cluster_4	-0.870	1.360	4.049	-0.065	-0.324	1.004	0.199	0.113	-3.0
2	2	cluster_0	-1.188	-0.733	-0.246	-0.065	-0.324	-0.757	0.199	0.113	0.32
3	3	cluster_0	-0.553	-0.733	-0.246	-1.371	-0.324	-0.757	0.199	0.113	0.32
4	4	cluster_0	-0.235	-0.733	-0.246	-1.371	-0.324	-0.757	0.199	0.113	0.32
5	5	cluster_0	-1.188	-0.733	-0.246	-0.065	-0.324	-0.757	0.199	0.113	0.32
6	6	cluster_4	-0.553	-0.733	-0.246	-0.065	-0.324	-0.757	0.199	0.113	-3.0
7	7	cluster_0	0.083	1.360	-0.246	-0.065	-0.324	1.004	0.199	0.113	0.32
8	8	cluster_0	-2.142	-0.733	-0.246	-0.065	-0.324	1.004	0.199	0.113	0.32
9	9	cluster_4	-1.188	-0.733	-0.246	1.892	-0.324	1.004	0.199	0.113	0.32
10	10	cluster_0	-0.235	-0.733	-0.246	-1.371	-0.324	1.004	0.199	0.113	0.32
11	11	cluster_0	0.401	-0.733	-0.246	-0.065	-0.324	1.004	0.199	0.113	0.32
12	12	cluster_0	0.719	1.360	-0.246	-0.065	-0.324	1.004	0.199	0.113	0.32
13	13	cluster_0	-0.870	-0.733	-0.246	-0.065	-0.324	-0.757	0.199	0.113	0.32

Figura 200. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.

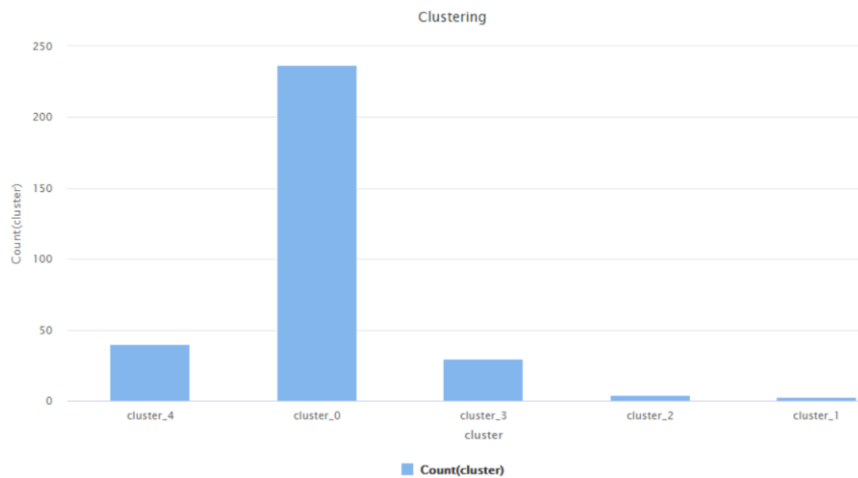


Figura 201. Representación gráfica de los clústers obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	ADMINIST...	ANALISIS ...	AUDITORI...	edad	genero	GESTION D...	provincia_...	sector_pr...	tipo_colegio	trabaja
Cluster 0	0.167	0.333	0.500	31.333	0.333	0.333	13.167	0	1.500	0
Cluster 1	1	0.818	1	31.364	0.091	0.909	10.091	0.091	1.636	1.000
Cluster 2	1	0.737	1	26.316	0.105	1	12.474	0.053	1	1
Cluster 3	1	0.939	0.994	28.673	0.436	0.982	10.982	0.061	1	-0
Cluster 4	1	0.920	1	28.850	0.292	0.965	11.310	0.053	2.106	0

Figura 202. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como Gestión de Redes, Auditoria Informática, Análisis Numérico y Administración de Centros de Cómputo, todas pertenecientes al octavo ciclo, si tiene una edad de 31 años, es de género masculino, oriundo de la provincia de Manabí, perteneciente al sector Urbano, proviene de un colegio Fiscal y no ejerce alguna actividad laboral.



**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo, todas pertenecientes al octavo ciclo, si tiene una edad de 31 años, es de género masculino, oriundo de la provincia de Imbabura, perteneciente al sector Urbano, proviene de un colegio Fiscomisional y ejerce alguna actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo, todas pertenecientes al octavo ciclo, si tiene una edad de 26 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscal y ejerce alguna actividad laboral.

**Clúster 3:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo, todas pertenecientes al octavo ciclo, si tiene una edad de 29 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal y no ejerce alguna actividad laboral.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Gestión de Redes, Auditoría Informática, Análisis Numérico y Administración de Centros de Cómputo, todas pertenecientes al octavo ciclo, si tiene una edad de 29 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional y no ejerce alguna actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide rojo y azul, se observa una estrecha relación entre los estudiantes de género masculino y un Colegio Fiscal y Fiscomisional, mientras que en el centroide de color verde hace referencia a estudiantes de género Femenino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, lo que indica que hay pocos estudiantes de la Carrera de Ingeniería en Sistemas que culminan sus estudios en instituciones privadas.

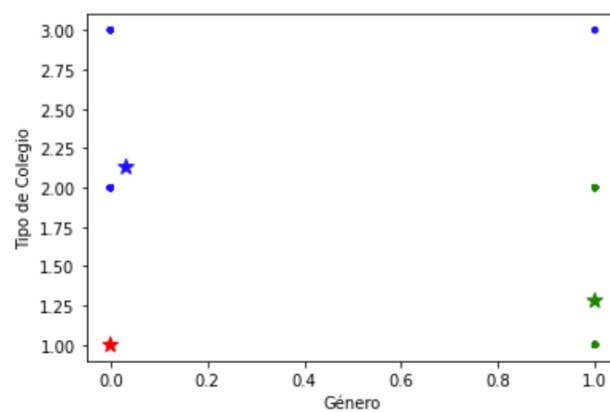


Figura 203. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Análisis Numérico** en donde los estudiantes aprueban a la asignatura si su edad es de 25, 29 y 34 años.

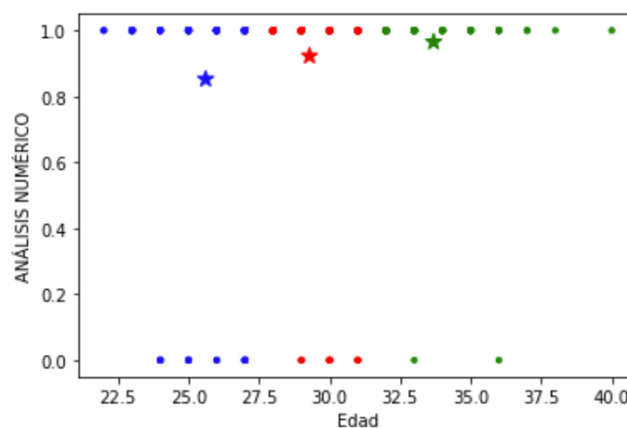


Figura 204. Relación Edad – Análisis Numérico

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Análisis numérico**, donde se observa que un grupo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprueban Análisis Numérico.

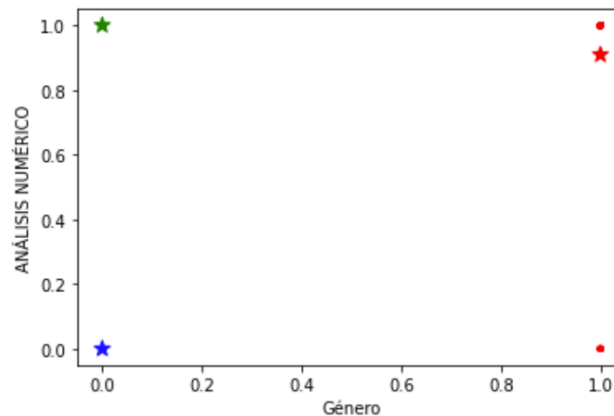


Figura 205. Relación Género – Análisis Numérico

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Gestión de Redes** e **Investigación de Operaciones**, en el clúster rojo se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde apruebe Gestión de Redes pero reprueba Investigación de Operaciones (clúster azul), y por último se muestra la probabilidad de que apruebe Investigación de Operaciones pero repruebe Gestión de Redes (clúster verde).

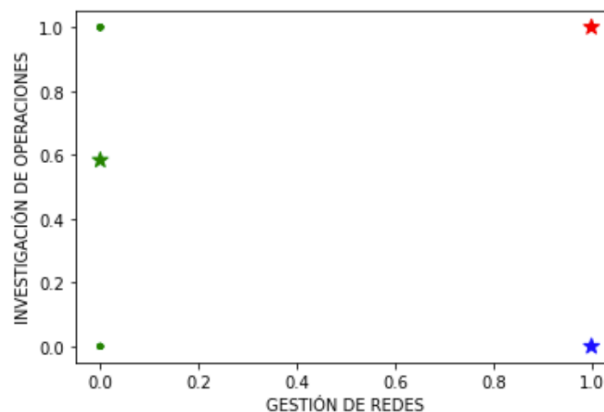


Figura 206. Relación Gestión de Redes – Investigación de Operaciones

## Anexo 12. Ingeniería en Sistemas – Noveno Ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **noveno ciclo de la carrera de Ingeniería en Sistemas** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas Weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

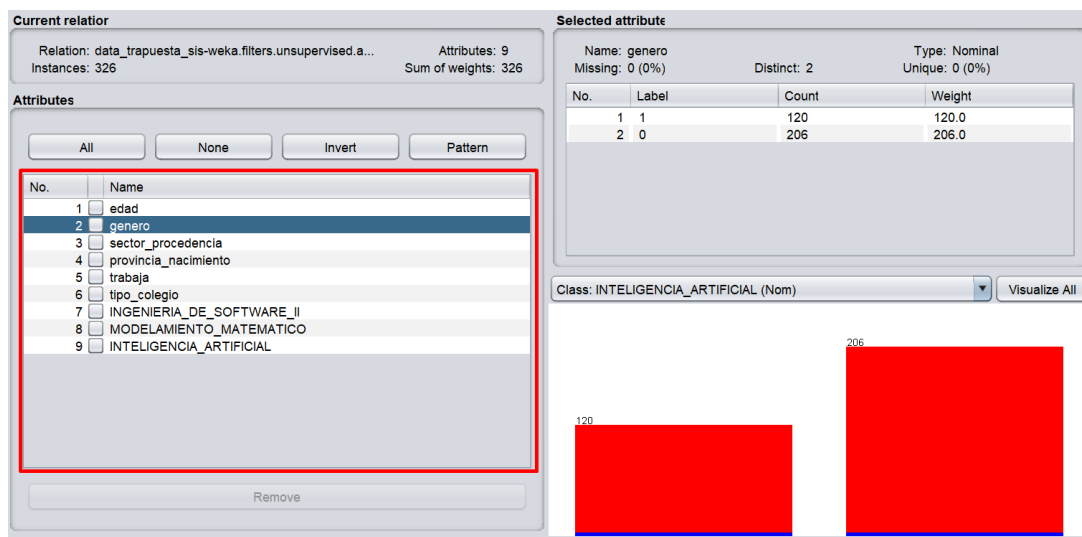


Figura 207. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.



```

Associator output
176. INGENIERIA_DE_SOFTWARE_II=1 MODELAMIENTO_MATEMATICO=1 310 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 273 <conf:(0.92)> lift:(1) lev:(0) [0]
177. trabaja=0 296 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 273 <conf:(0.92)> lift:(1) lev:(0) [0]
178. trabaja=0 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 282 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 293 <conf:(0.92)> lift:(1) lev:(0) [0]
179. trabaja=0 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 282 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 318 <conf:(0.92)> lift:(1) lev:(0) [0]
180. MODELAMIENTO_MATEMATICO=1 318 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 293 <conf:(0.92)> lift:(1) lev:(0) [0]
181. MODELAMIENTO_MATEMATICO=1 318 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 293 <conf:(0.92)> lift:(1) lev:(0) [0]
182. INGENIERIA_DE_SOFTWARE_II=1 317 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 292 <conf:(0.92)> lift:(1) lev:(0) [0]
183. provincia_nacimiento=11 INTELIGENCIA_ARTIFICIAL=1 252 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 292 <conf:(0.92)> lift:(1) lev:(0) [0]
184. MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 311 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 292 <conf:(0.92)> lift:(1) lev:(0) [0]
185. provincia_nacimiento=11 trabaja=0 235 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 216 <conf:(0.92)> lift:(1) lev:(0) [0]
186. trabaja=0 296 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 272 <conf:(0.92)> lift:(1) lev:(0) [0]
187. trabaja=0 296 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 272 <conf:(0.92)> lift:(1) lev:(0) [0]
188. provincia_nacimiento=11 259 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 238 <conf:(0.92)> lift:(1) lev:(0) [0]
189. INTELIGENCIA_ARTIFICIAL=1 319 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 293 <conf:(0.92)> lift:(1) lev:(0) [0]
190. trabaja=0 INTELIGENCIA_ARTIFICIAL=1 290 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 266 <conf:(0.92)> lift:(1) lev:(0) [0]
191. trabaja=0 INTELIGENCIA_ARTIFICIAL=1 290 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 266 <conf:(0.92)> lift:(1) lev:(0) [0]
192. provincia_nacimiento=11 INTELIGENCIA_ARTIFICIAL=1 252 ==> sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 238 <conf:(0.92)> lift:(1) lev:(0) [0]
193. provincia_nacimiento=11 INGENIERIA_DE_SOFTWARE_II=1 MODELAMIENTO_MATEMATICO=1 249 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 252 <conf:(0.92)> lift:(1) lev:(0) [0]
194. INTELIGENCIA_ARTIFICIAL=1 319 ==> sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 292 <conf:(0.92)> lift:(1) lev:(0) [0]
195. provincia_nacimiento=11 trabaja=0 235 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 216 <conf:(0.92)> lift:(1) lev:(0) [0]
196. provincia_nacimiento=11 INGENIERIA_DE_SOFTWARE_II=1 254 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 216 <conf:(0.92)> lift:(1) lev:(0) [0]
197. provincia_nacimiento=11 MODELAMIENTO_MATEMATICO=1 253 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 236 <conf:(0.92)> lift:(1) lev:(0) [0]
198. provincia_nacimiento=11 259 ==> sector_procedencia=0 INTELIGENCIA_ARTIFICIAL=1 236 <conf:(0.92)> lift:(1) lev:(0) [0]
199. sector_procedencia=0 provincia_nacimiento=11 INTELIGENCIA_ARTIFICIAL=1 236 ==> trabaja=0 215 <conf:(0.92)> lift:(1) lev:(0) [0]
200. sector_procedencia=0 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 292 ==> trabaja=0 266 <conf:(0.92)> lift:(1) lev:(0) [0]
  
```

Figura 208. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados. Aplicado los filtros para reglas con nivel de confianza mayor a 98% y lift mayor a 1 respectivamente, se obtiene un total de 19 reglas (Ver Figura 3).

REGLAS	CONF.	LIFT.	LEV.	
1. sector_procedencia=0 provincia_nacimiento=11 MODELAMIENTO_MATEMATICO=1 238 ==> INGENIERIA_DE_SOFTWARE_II=1 235	conf:(0.99)	lift:(1.02)	lev:(0.01)	[3]
2. sector_procedencia=0 provincia_nacimiento=11 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 231 ==> INGENIERIA_DE_SOFTWARE_II=1 228	conf:(0.99)	lift:(1.02)	lev:(0.01)	[3]
3. sector_procedencia=0 provincia_nacimiento=11 trabaja=0 MODELAMIENTO_MATEMATICO=1 216 ==> INGENIERIA_DE_SOFTWARE_II=1 213	conf:(0.99)	lift:(1.01)	lev:(0.01)	[2]
4. provincia_nacimiento=11 MODELAMIENTO_MATEMATICO=1 253 ==> INGENIERIA_DE_SOFTWARE_II=1 249	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
5. provincia_nacimiento=11 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 246 ==> INGENIERIA_DE_SOFTWARE_II=1 242	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
6. sector_procedencia=0 provincia_nacimiento=11 243 ==> INGENIERIA_DE_SOFTWARE_II=1 239	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
7. sector_procedencia=0 provincia_nacimiento=11 INGENIERIA_DE_SOFTWARE_II=1 239 ==> MODELAMIENTO_MATEMATICO=1 235	conf:(0.98)	lift:(1.01)	lev:(0.01)	[1]
8. sector_procedencia=0 provincia_nacimiento=11 INTELIGENCIA_ARTIFICIAL=1 236 ==> INGENIERIA_DE_SOFTWARE_II=1 232	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
9. sector_procedencia=0 provincia_nacimiento=11 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 232 ==> MODELAMIENTO_MATEMATICO=1 228	conf:(0.98)	lift:(1.01)	lev:(0.01)	[1]
10. provincia_nacimiento=11 trabaja=0 MODELAMIENTO_MATEMATICO=1 229 ==> INGENIERIA_DE_SOFTWARE_II=1 225	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
11. provincia_nacimiento=11 trabaja=0 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 223 ==> INGENIERIA_DE_SOFTWARE_II=1 219	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
12. sector_procedencia=0 provincia_nacimiento=11 trabaja=0 221 ==> INGENIERIA_DE_SOFTWARE_II=1 217	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
13. sector_procedencia=0 provincia_nacimiento=11 trabaja=0 INGENIERIA_DE_SOFTWARE_II=1 217 ==> MODELAMIENTO_MATEMATICO=1 213	conf:(0.98)	lift:(1.01)	lev:(0)	[1]
14. provincia_nacimiento=11 259 ==> INGENIERIA_DE_SOFTWARE_II=1 254	conf:(0.98)	lift:(1.01)	lev:(0.01)	[2]
16. provincia_nacimiento=11 INTELIGENCIA_ARTIFICIAL=1 252 ==> INGENIERIA_DE_SOFTWARE_II=1 247	conf:(0.98)	lift:(1.01)	lev:(0.01)	[1]
25. provincia_nacimiento=11 trabaja=0 235 ==> INGENIERIA_DE_SOFTWARE_II=1 230	conf:(0.98)	lift:(1.01)	lev:(0)	[1]
29. provincia_nacimiento=11 trabaja=0 INTELIGENCIA_ARTIFICIAL=1 229 ==> INGENIERIA_DE_SOFTWARE_II=1 224	conf:(0.98)	lift:(1.01)	lev:(0)	[1]
31. sector_procedencia=0 trabaja=0 MODELAMIENTO_MATEMATICO=1 272 ==> INGENIERIA_DE_SOFTWARE_II=1 266	conf:(0.98)	lift:(1.01)	lev:(0)	[1]
38. sector_procedencia=0 trabaja=0 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 266 ==> INGENIERIA_DE_SOFTWARE_II=1 260	conf:(0.98)	lift:(1.01)	lev:(0)	[1]

Figura 209. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:





UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

- 1) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y aprueba Modelamiento Matemático es muy probable que también apruebe la asignatura Ingeniería del software II.
- 2) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y aprueba Modelamiento Matemático e Inteligencia Artificial, es muy probable que también apruebe la asignatura Ingeniería del software II.
- 3) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja, no ejerce actividad laboral y aprueba Modelamiento Matemático es muy probable que también apruebe la asignatura Ingeniería del software II.
- 4) Si un estudiante pertenece a la provincia de Loja y aprueba Modelamiento Matemático es muy probable que también apruebe la asignatura Ingeniería del software II.
- 5) Si un estudiante pertenece a la provincia de Loja y aprueba Modelamiento Matemático e Inteligencia Artificial, es muy probable que también apruebe la asignatura Ingeniería del software II.
- 6) Si un estudiante es procedente del sector Urbano y pertenece a la provincia de Loja es muy probable que también apruebe la asignatura Ingeniería del software II.
- 7) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y aprueba Ingeniería del Software, es muy probable que también apruebe la asignatura Modelamiento Matemático.
- 8) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y aprueba Ingeniería del Software e Inteligencia Artificial, es muy probable que también apruebe la asignatura Modelamiento Matemático.
- 9) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad académica y aprueba Modelamiento Matemático, es muy probable que también apruebe la asignatura de Ingeniería de Software II.
- 10) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad académica y aprueba Modelamiento Matemático e Inteligencia Artificial, es muy probable que también apruebe la asignatura de Ingeniería de Software II.

- 11) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja y no ejerce actividad laboral es muy probable que también apruebe la asignatura Ingeniería del software II.
- 12) Si un estudiante es procedente del sector Urbano, pertenece a la provincia de Loja, no ejerce actividad laboral y aprueba la asignatura de Ingeniería de Software II, es muy probable que también apruebe la asignatura Modelamiento Matemático.
- 13) Si un estudiante es procedente del sector Urbano es muy probable que apruebe la Ingeniería del Software II.
- 14) Si un estudiante es procedente del sector Urbano y aprueba Inteligencia Artificial, es muy probable que apruebe la Ingeniería del Software II.
- 15) Si un estudiante es procedente del sector Urbano y no ejerce una actividad laboral, es muy probable que apruebe la Ingeniería del Software II.
- 16) Si un estudiante pertenece a la provincia de Loja, no ejerce actividad académica y aprueba Inteligencia Artificial, es muy probable que también apruebe la asignatura de Ingeniería de Software II.
- 17) Si un estudiante es procedente del sector Urbano y aprueba Modelamiento Matemático es muy probable que también apruebe la asignatura Ingeniería del software II.
- 18) Si un estudiante es procedente del sector Urbano y aprueba Modelamiento Matemático e Inteligencia Artificial, es muy probable que también apruebe la asignatura Ingeniería del software II.

Por otra parte, se han determinado Reglas definiendo Lift como predeterminado, obteniendo valores mayores a 1.18 y una confiabilidad de las premisas del 85% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CON.	LIFT.	
12. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 100 ==> provincia_nacimiento=11 trabaja=0 MODELAMIENTO_MATEMATICO=1 86	0.86)	lift:(1.22)>	lev:(0.05)
33. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 102 ==> provincia_nacimiento=11 trabaja=0 MODELAMIENTO_MATEMATICO=1 87 conf:	0.85)	lift:(1.21)>	lev:(0.05)
64. sector_procedencia=0 tipo_colegio=2 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 101 ==> provincia_nacimiento=11 trabaja=0 INGENIERIA_DE_SOFTWARE_II=1 86	0.85)	lift:(1.21)>	lev:(0.05)
105. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 100 ==> provincia_nacimiento=11 trabaja=0 86 conf:	0.86)	lift:(1.19)>	lev:(0.04)
106. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 100 ==> provincia_nacimiento=11 trabaja=0 86	0.86)	lift:(1.19)>	lev:(0.04)
153. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 102 ==> provincia_nacimiento=11 trabaja=0 87 conf:	0.85)	lift:(1.18)>	lev:(0.04)
154. sector_procedencia=0 tipo_colegio=2 INTELIGENCIA_ARTIFICIAL=1 102 ==> provincia_nacimiento=11 trabaja=0 87 conf:	0.85)	lift:(1.18)>	lev:(0.04)
155. sector_procedencia=0 tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 MODELAMIENTO_MATEMATICO=1 102 ==> provincia_nacimiento=11 trabaja=0 87 conf:	0.85)	lift:(1.18)>	lev:(0.04)
164. tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 INTELIGENCIA_ARTIFICIAL=1 108 ==> provincia_nacimiento=11 trabaja=0 92 conf:	0.85)	lift:(1.18)>	lev:(0.04)
169. sector_procedencia=0 tipo_colegio=2 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 101 ==> provincia_nacimiento=11 trabaja=0 86 conf:	0.85)	lift:(1.18)>	lev:(0.04)
180. tipo_colegio=2 INGENIERIA_DE_SOFTWARE_II=1 MODELAMIENTO_MATEMATICO=1 INTELIGENCIA_ARTIFICIAL=1 107 ==> provincia_nacimiento=11 trabaja=0 91 conf:	0.85)	lift:(1.18)>	lev:(0.04)
189. genero=1 sector_procedencia=0 MODELAMIENTO_MATEMATICO=1 113 ==> provincia_nacimiento=11 trabaja=0 96 conf:	0.85)	lift:(1.18)>	lev:(0.04)

Figura 210. Reglas obtenidas en Weka con Lift mayor a 1



- 1) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II e Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja, no ejerza actividad laboral y apruebe la asignatura de Modelamiento Matemático.
- 2) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II es muy probable que pertenezca a la provincia de Loja, no ejerza actividad laboral y apruebe la asignatura de Modelamiento Matemático.
- 3) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Modelamiento Matemático e Inteligencia Artificial es muy probable que pertenezca a la provincia de Loja, no ejerza actividad laboral y apruebe la asignatura de Ingeniería de Software II.
- 4) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II, es muy probable que pertenezca a la provincia de Loja y no ejerce actividad laboral.
- 5) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja y no ejerce actividad laboral.
- 6) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.
- 7) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.
- 8) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba

Ingeniería de Software II y Modelamiento Matemático, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.

- 9) Si un estudiante ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II e Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.
- 10) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Modelamiento Matemático e Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.
- 11) Si un estudiante es procedente del sector Urbano, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.
- 12) Si un estudiante es de género femenino, procedente del sector Urbano y aprueba Modelamiento Matemático, es muy probable que pertenezca a la provincia de Loja y no ejerza actividad laboral.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es bajo, pero ya se demuestra la presencia de mujeres en ciclos superiores como este caso.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, se muestra la presencia de estudiante provenientes de la Provincia de Loja, además, en este caso se muestra la característica tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales,

Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 32 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

Association Rules

```
[provincia_nacimiento] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.867)
[provincia_nacimiento] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.867)
[INTELIGENCIA ARTIFICIAL, provincia_nacimiento] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.867)
[sector_procedencia] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.895)
[sector_procedencia] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.895)
[INTELIGENCIA ARTIFICIAL, sector_procedencia] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.895)
[MODELAMIENTO MATEMATICO, provincia_nacimiento] --> [INGENIERIA DE SOFTWARE II] (confidence: 0.897)
[MODELAMIENTO MATEMATICO, provincia_nacimiento] --> [INTELIGENCIA ARTIFICIAL, INGENIERIA DE SOFTWARE II] (confidence: 0.897)
[INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, provincia_nacimiento] --> [INGENIERIA DE SOFTWARE II] (confidence: 0.897)
[provincia_nacimiento] --> [INGENIERIA DE SOFTWARE II] (confidence: 0.900)
[provincia_nacimiento] --> [INTELIGENCIA ARTIFICIAL, INGENIERIA DE SOFTWARE II] (confidence: 0.900)
[INTELIGENCIA ARTIFICIAL, provincia_nacimiento] --> [INGENIERIA DE SOFTWARE II] (confidence: 0.900)
[edad, tipo_colegio] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.909)
[edad, genero] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO] (confidence: 0.915)
[edad, genero] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.915)
[INGENIERIA DE SOFTWARE II, edad, genero] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO] (confidence: 0.915)
[edad] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.915)
[INTELIGENCIA ARTIFICIAL, edad, tipo_colegio] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.923)
[edad, tipo_colegio] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.924)
[INTELIGENCIA ARTIFICIAL, edad] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.929)
[edad] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO] (confidence: 0.930)
[edad] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.930)
[edad] --> [INTELIGENCIA ARTIFICIAL, MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.930)
[INTELIGENCIA ARTIFICIAL, edad, genero] --> [MODELAMIENTO MATEMATICO] (confidence: 0.931)
[INTELIGENCIA ARTIFICIAL, edad, genero] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.931)
[INTELIGENCIA ARTIFICIAL, INGENIERIA DE SOFTWARE II, edad, genero] --> [MODELAMIENTO MATEMATICO] (confidence: 0.931)
[edad, genero] --> [MODELAMIENTO MATEMATICO] (confidence: 0.932)
[edad, genero] --> [MODELAMIENTO MATEMATICO, INGENIERIA DE SOFTWARE II] (confidence: 0.932)
```

Figura 211. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante pertenece al sector Urbano y es oriundo de la provincia de Loja, estas son las características más relevantes que se destacan en las premisas, por otro lado, se identifica la aprobación de materias en forma paralela en conjunto con lo anteriormente mencionado.



```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Guardando...
Comentar Compartir

+ Código + Texto
RAM Disco Editando
RelationRecord(items=frozenset({'IS_A', '11', 'MM_A'}), support=0.7638036809815951, ordered_statistics=[OrderedStatist
RelationRecord(items=frozenset({'IS_A', 'Urbano', '11'}), support=0.7331288343558282, ordered_statistics=[OrderedStat
RelationRecord(items=frozenset({'IS_A', 'IA_A', '11', 'MM_A'}), support=0.7423312883435583, ordered_statistics=[Order
RelationRecord(items=frozenset({'IS_A', 'Urbano', 'IA_A', '11'}), support=0.7116564417177914, ordered_statistics=[Ord
RelationRecord(items=frozenset({'IS_A', 'Urbano', '11', 'MM_A'}), support=0.7208588957055214, ordered_statistics=[Ord

```

Figura 212. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

#### Clustered Instances

0	142 ( 44%)
1	78 ( 24%)
2	76 ( 23%)
3	12 ( 4%)
4	18 ( 6%)

Figura 213. Clúster con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 1, y 2. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Full Data (326.0)	Cluster#				
		0 (142.0)	1 (78.0)	2 (76.0)	3 (12.0)	4 (18.0)
edad	29	29	32	30	27	25
genero	0	0	1	0	1	0
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	0	0	1
tipo_colegio	1	1	1	2	1	1
INGENIERIA_DE_SOFTWARE_II	1	1	1	1	1	1
MODELAMIENTO_MATEMATICO	1	1	1	1	1	1
INTELIGENCIA_ARTIFICIAL	1	1	1	1	1	1

Figura 214. Clústeres obtenidos por Weka.

**Clúster\_0 (44%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja,



que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al noveno ciclo, estas son: Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial.

**Clúster\_1 (24%):** Estudiantes que tengan una edad de 32 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al noveno ciclo, estas son: Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial.

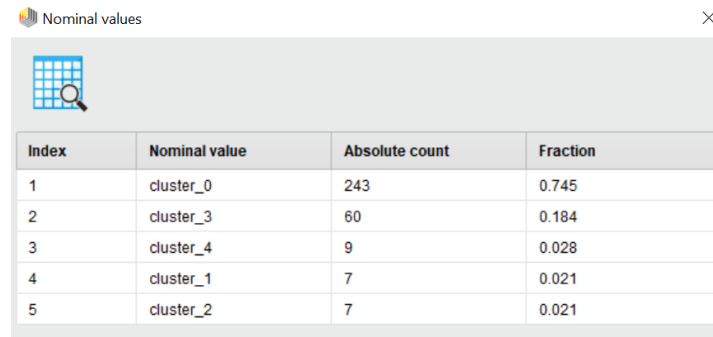
**Clúster\_2 (23%):** Estudiantes que tengan una edad de 30 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al noveno ciclo, estas son: Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial.

**Clúster\_3 (4%):** Estudiantes que tengan una edad de 27 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al noveno ciclo, estas son: Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial.

**Clúster\_4 (6%):** Estudiantes que tengan una edad de 25 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que ejerzan una actividad laboral y que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al noveno ciclo, estas son: Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial.

## Resultados obtenidos por Algoritmo K - means – RapidMiner

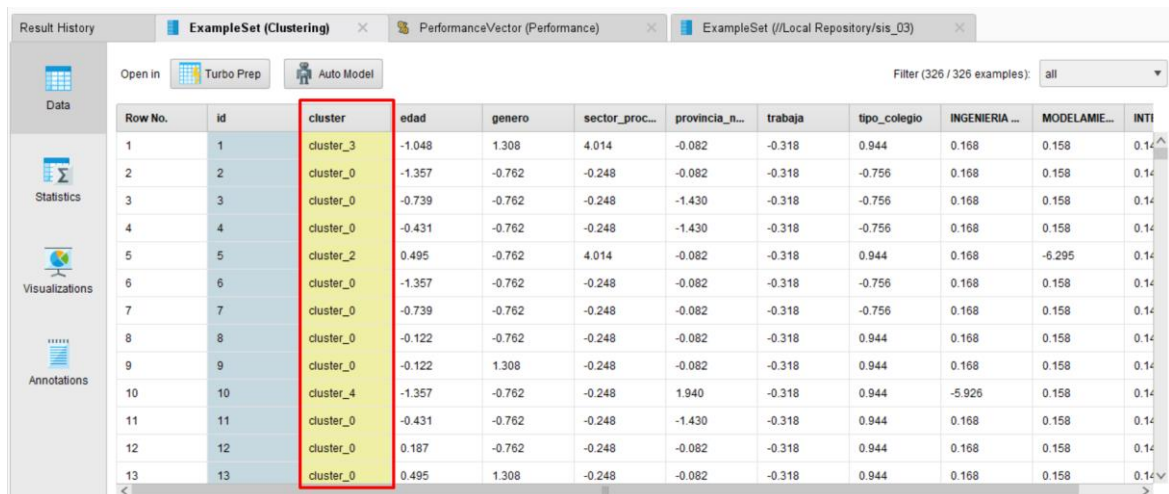
La Figura 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	243	0.745
2	cluster_3	60	0.184
3	cluster_4	9	0.028
4	cluster_1	7	0.021
5	cluster_2	7	0.021

Figura 215. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 326 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	INGENIERIA ...	MODELAMIE...	INTI
1	1	cluster_3	-1.048	1.308	4.014	-0.082	-0.318	0.944	0.168	0.158	0.14
2	2	cluster_0	-1.357	-0.762	-0.248	-0.082	-0.318	-0.756	0.168	0.158	0.14
3	3	cluster_0	-0.739	-0.762	-0.248	-1.430	-0.318	-0.756	0.168	0.158	0.14
4	4	cluster_0	-0.431	-0.762	-0.248	-1.430	-0.318	-0.756	0.168	0.158	0.14
5	5	cluster_2	0.495	-0.762	4.014	-0.082	-0.318	0.944	0.168	-6.295	0.14
6	6	cluster_0	-1.357	-0.762	-0.248	-0.082	-0.318	-0.756	0.168	0.158	0.14
7	7	cluster_0	-0.739	-0.762	-0.248	-0.082	-0.318	-0.756	0.168	0.158	0.14
8	8	cluster_0	-0.122	-0.762	-0.248	-0.082	-0.318	0.944	0.168	0.158	0.14
9	9	cluster_0	-0.122	1.308	-0.248	-0.082	-0.318	0.944	0.168	0.158	0.14
10	10	cluster_4	-1.357	-0.762	-0.248	1.940	-0.318	0.944	-5.925	0.158	0.14
11	11	cluster_0	-0.431	-0.762	-0.248	-1.430	-0.318	0.944	0.168	0.158	0.14
12	12	cluster_0	0.187	-0.762	-0.248	-0.082	-0.318	0.944	0.168	0.158	0.14
13	13	cluster_0	0.495	1.308	-0.248	-0.082	-0.318	0.944	0.168	0.158	0.14

Figura 216. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.



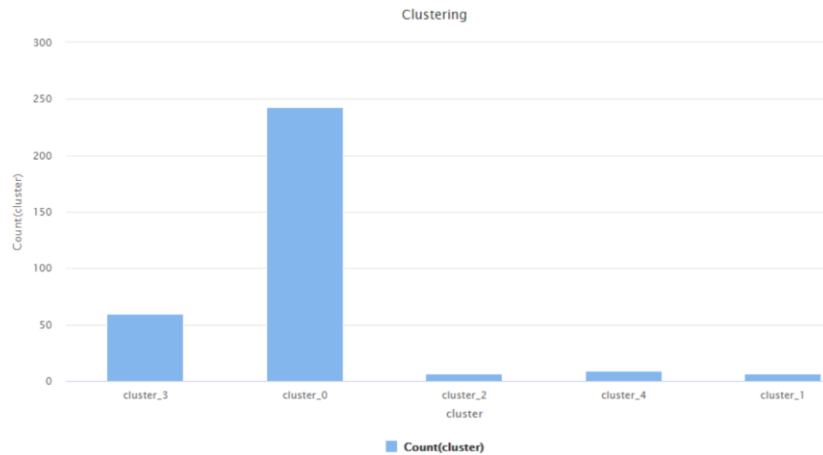


Figura 217. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	edad	genero	INGENIERIA ...	INTELIGENCI...	MODELAMIE...	provincia_n...	sector_proc...	tipo_colegio	trabaja
Cluster 0	28.718	0.017	1	0.974	1	11.009	0.077	1.060	0.239
Cluster 1	29.687	0.036	1	0.988	1	11.301	0.060	2.169	-0
Cluster 2	29.736	1	1	0.973	1	11.300	0.027	1.300	0.009
Cluster 3	27.000	0.111	0	1	0.889	12.778	0.111	1.556	0.111
Cluster 4	35.000	0.571	1	1	0	11.571	0.143	1.429	0

Figura 218. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial, todas pertenecientes al noveno ciclo, si tiene una edad de 28 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 1:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial, todas pertenecientes al noveno ciclo, si tiene una edad de



29 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y no ejerce alguna actividad laboral.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Ingeniería de Software II, Modelamiento Matemático e Inteligencia Artificial, todas pertenecientes al noveno ciclo, si tiene una edad de 30 años, es de género femenino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y no ejerce alguna actividad laboral.

**Clúster 3:** Estudiante aprueba dos asignaturas relacionadas a las matemáticas como Modelamiento Matemático e Inteligencia Artificial, mientras que, reprueba la materia de Ingeniería de Software II, todas pertenecientes al noveno ciclo, si tiene una edad de 27 años, es de género masculino, oriundo de la provincia de Los Ríos, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y no ejerce alguna actividad laboral.

**Clúster 0:** Estudiante aprueba dos asignaturas relacionadas a las matemáticas como Ingeniería de Software II e Inteligencia Artificial y reprueba Modelamiento Matemático, todas pertenecientes al noveno ciclo, si tiene una edad de 35 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y a parte de sus actividades académicas ejerce alguna actividad laboral.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con

sus respectivos centroides; considerado al centroide rojo y verde, se observa una relación entre los estudiantes de género masculino y una Institución Fiscal y Fiscomisional, mientras que en el centroide de color azul hace referencia a estudiantes de género Femenino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, lo que indica que hay pocos estudiantes de la Carrera de Ingeniería en Sistemas que culminan sus estudios en instituciones privadas.

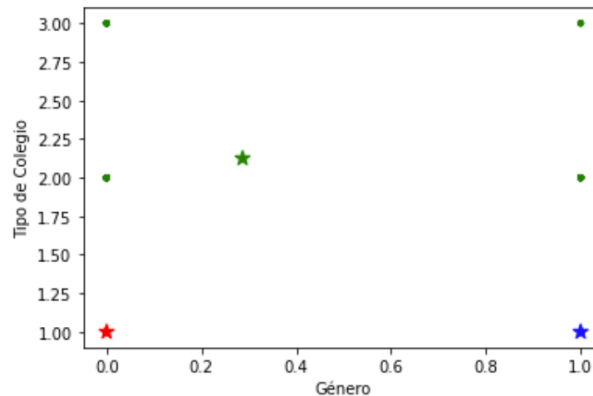


Figura 219. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Ingeniería de Software II** en donde los estudiantes aprueban a la asignatura si su edad es de 26, 30 y 34 años.

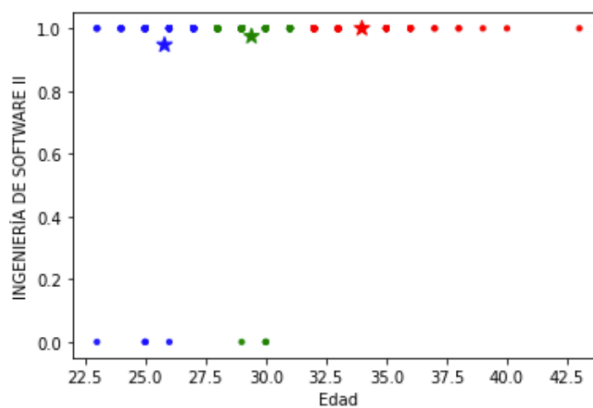


Figura 220. Relación Edad – Ingeniería de Software II

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** e **Ingeniería de Software II**, donde se observa que un grupo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de

estudiantes pertenecientes al mismo género que reprueban la asignatura de Ingeniería de Software II.

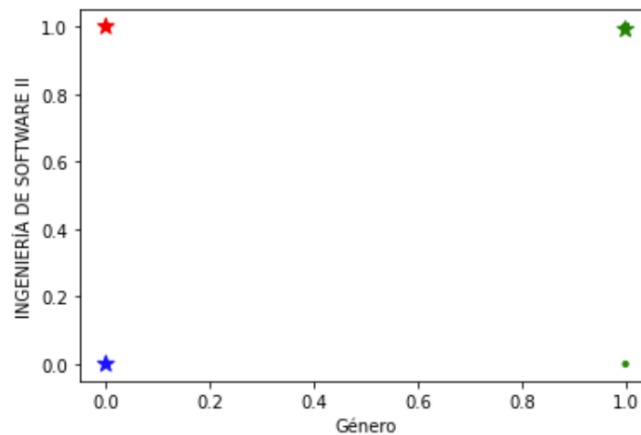


Figura 221. Relación Género – Ingeniería de Software II

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Modelamiento Matemático** e **Ingeniería de Software II**, en el clúster rojo se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde apruebe Modelamiento Matemático pero reprueba Ingeniería de Software II (clúster azul), y por último se muestra una alta probabilidad que el estudiante únicamente apruebe Ingeniería de Software II (clúster verde).

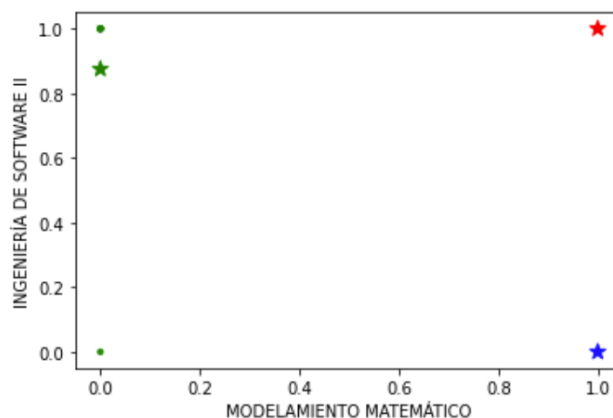


Figura 222. Relación Modelamiento Matemático – Ingeniería de Software II

### Anexo 13. Ingeniería en Sistemas – Décimo ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

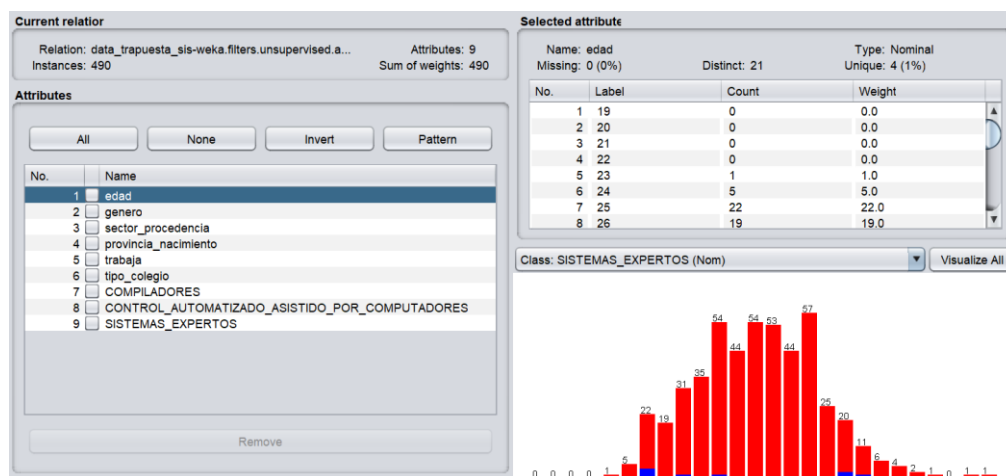


Figura 223. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
177. sector_procedencia=0 470 ==> trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 433 <conf:(0.92)>
178. sector_procedencia=0 470 ==> COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPER
179. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 475 ==> sector_procedencia=0 COMPILADORES=1 437 <conf:(
180. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 471 ==> sector_procedencia=0 COMPILADO
181. trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 452 ==> sector_procedencia=0 COMPILADORES=1 415
182. provincia_nacimiento=11 402 ==> trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTO
183. trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 449 ==> sector_procedencia=0
184. SISTEMAS_EXPERTOS=1 482 ==> sector_procedencia=0 COMPILADORES=1 442 <conf:(0.92)> lift:(1.01) lev:(0.01)
185. sector_procedencia=0 470 ==> trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1
186. trabaja=0 SISTEMAS_EXPERTOS=1 457 ==> sector_procedencia=0 COMPILADORES=1 418 <conf:(0.91)> lift:(1) lev
187. sector_procedencia=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 451 ==> trabaja=0
188. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 471 ==> sector_procedencia=0 trabaja=0
189. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 471 ==> trabaja=0 COMPILADORES=1 430
190. sector_procedencia=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 455 ==> trabaja=0 COMPILADORES=1 415
191. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 475 ==> sector_procedencia=0 trabaja=0 433 <conf:(0.91)
192. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 475 ==> trabaja=0 COMPILADORES=1 433 <conf:(0.91)> lift
193. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 475 ==> sector_procedencia=0 COMPILADORES=1 SISTEMAS_EXPER
194. COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 452 ==> sector_proceden
195. trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 452 ==> sector_procedencia=0 COMPILADORES=1 SIST
196. COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 456 ==> sector_procedencia=0 trabaja=0 415
197. SISTEMAS_EXPERTOS=1 482 ==> sector_procedencia=0 trabaja=0 438 <conf:(0.91)> lift:(1) lev:(0) [0] conv:(
198. trabaja=0 464 ==> sector_procedencia=0 COMPILADORES=1 421 <conf:(0.91)> lift:(1) lev:(-0) [-1] conv:(0.9
199. COMPILADORES=1 SISTEMAS_EXPERTOS=1 461 ==> sector_procedencia=0 trabaja=0 418 <conf:(0.91)> lift:(1) lev
200. sector_procedencia=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 455 ==> trabaja=0 COMPILADORES=1 SIST

```

Figura 224. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza de 0.99 y lift mayor a 1 respectivamente, se obtiene un total de 17 reglas (Ver Figura 3).

REGLAS	CONFIANZA	LIFT	
1. provincia_nacimiento=11 trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 371 ==> SISTEMAS_EXPERTOS=1 369	conf:(0.99)>	ift:(1.01) lev:(0.01)	[4]
2. trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 452 ==> SISTEMAS_EXPERTOS=1 449	conf:(0.99)>	ift:(1.01) lev:(0.01)	[4]
3. trabaja=0 COMPILADORES=1 439 ==> SISTEMAS_EXPERTOS=1 436	conf:(0.99)>	ift:(1.01) lev:(0.01)	[4]
4. sector_procedencia=0 trabaja=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 433 ==> SISTEMAS_EXPERTOS=1 430	conf:(0.99)>	ift:(1.01) lev:(0.01)	[4]
5. trabaja=0 COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 433 ==> SISTEMAS_EXPERTOS=1 430	conf:(0.99)>	ift:(1.01) lev:(0.01)	[4]
6. sector_procedencia=0 trabaja=0 COMPILADORES=1 421 ==> SISTEMAS_EXPERTOS=1 418	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
7. sector_procedencia=0 trabaja=0 COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 415 ==> SISTEMAS_EXPERTOS=1 387	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
8. provincia_nacimiento=11 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 390 ==> SISTEMAS_EXPERTOS=1 387	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
9. provincia_nacimiento=11 COMPILADORES=1 381 ==> SISTEMAS_EXPERTOS=1 378	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
10. sector_procedencia=0 provincia_nacimiento=11 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 374 ==> SISTEMAS_EXPERTOS=1 374	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
11. provincia_nacimiento=11 COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 374 ==> SISTEMAS_EXPERTOS=1 374	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
12. CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 475 ==> SISTEMAS_EXPERTOS=1 471	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
13. COMPILADORES=1 465 ==> SISTEMAS_EXPERTOS=1 461	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
14. COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 456 ==> SISTEMAS_EXPERTOS=1 452	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
15. sector_procedencia=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 455 ==> SISTEMAS_EXPERTOS=1 451	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
16. sector_procedencia=0 COMPILADORES=1 446 ==> SISTEMAS_EXPERTOS=1 442	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]
17. sector_procedencia=0 COMPILADORES=1 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 437 ==> SISTEMAS_EXPERTOS=1 433	conf:(0.99)>	ift:(1.01) lev:(0.01)	[3]

Figura 225. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante pertenece a la ciudad de Loja, no ejerce ninguna actividad laboral y aprueba la materia de Control Automatizado Asistido





- por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
- 2) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 3) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Compiladores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 4) Si un estudiante pertenece al sector Urbano, no ejerce ninguna actividad laboral y aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 5) Si un estudiante no ejerce ninguna actividad laboral y aprueba Compiladores y Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 6) Si un estudiante pertenece al sector Urbano, no ejerce ninguna actividad laboral y aprueba la materia de Compiladores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 7) Si un estudiante pertenece al sector Urbano, no ejerce ninguna actividad laboral y aprueba a Compiladores y Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 8) Si un estudiante pertenece a la ciudad de Loja y aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 9) Si un estudiante pertenece a la ciudad de Loja y aprueba la materia de Compiladores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 10) Si un estudiante pertenece al sector Urbano y es oriundo de la ciudad de Loja, no ejerce ninguna actividad laboral y aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
  - 11) Si un estudiante es oriundo de la provincia de Loja y aprueba las asignaturas de Compiladores y Control Automatizado Asistido por

Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.

- 12) Si un estudiante aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
- 13) Si un estudiante aprueba la materia de Compiladores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
- 14) Si un estudiante aprueba a las asignaturas de Compiladores y Control Automatizado Asistido por Computadores, es muy probable que apruebe también la materia de Sistemas Expertos.
- 15) Si un estudiante pertenece al sector Urbano y aprueba la materia de Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
- 16) Si un estudiante pertenece al sector Urbano y aprueba a Compiladores, es muy probable que apruebe también la asignatura de Sistemas Expertos.
- 17) Si un estudiante pertenece al sector Urbano y aprueba a Compiladores y Control Automatizado Asistido por Computadores, es muy probable que apruebe también la asignatura de Sistemas Expertos.

Por otra parte, se han determinado Reglas definiendo a lift con su valor más elevado, en este caso únicamente se obtienen valores del 1.02 y una confianza mayor al 95% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONFIANZA	LIFT	
3. trabaja=0 COMPILADORES=1 439 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXPERTOS=1 430 con	:(0.98)	lift:(1.02)>	ev:(0.02) [8]
5. sector_procedencia=0 trabaja=0 COMPILADORES=1 421 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 SISTEMAS_EXP	:(0.98)	lift:(1.02)>	ev:(0.01) [7]
7. trabaja=0 COMPILADORES=1 439 ==> sector_procedencia=0 CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 415 con	:(0.95)	lift:(1.02)>	ev:(0.02) [7]
11. trabaja=0 COMPILADORES=1 439 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 433 con	:(0.99)	lift:(1.02)>	ev:(0.02) [7]
14. trabaja=0 COMPILADORES=1 SISTEMAS_EXPERTOS=1 436 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 430 con	:(0.99)	lift:(1.02)>	ev:(0.01) [7]
16. sector_procedencia=0 trabaja=0 COMPILADORES=1 421 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES=1 415 con	:(0.99)	lift:(1.02)>	ev:(0.01) [6]
18. sector_procedencia=0 trabaja=0 COMPILADORES=1 SISTEMAS_EXPERTOS=1 418 ==> CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTA	:(0.99)	lift:(1.02)>	ev:(0.01) [6]

Figura 226. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante no trabaja y aprueba la asignatura de Compiladores, es muy probable que también apruebe Control Automatizado Asistido por Computadores y Sistemas Expertos.
- 2) Si un estudiante es del sector Urbano, no ejerce actividad laboral y aprueba la asignatura de Compiladores, es muy probable que también



- apruebe Control Automatizado Asistido por Computadores y Sistemas Expertos.
- 3) Si un estudiante no ejerce actividad laboral y aprueba la asignatura de Compiladores, es muy probable que pertenezca al sector Urbano y también apruebe Control Automatizado Asistido por Computadores.
  - 4) Si un estudiante no ejerce actividad laboral y aprueba la asignatura de Compiladores, es muy probable que también apruebe Control Automatizado Asistido por Computadores.
  - 5) Si un estudiante no ejerce actividad laboral y aprueba la asignatura de Compiladores y Sistemas Expertos, es muy probable que apruebe Control Automatizado Asistido por Computadores.
  - 6) Si un estudiante es del sector Urbano, no ejerce actividad laboral y aprueba la asignatura de Compiladores, es muy probable que también apruebe Control Automatizado Asistido por Computadores.
  - 7) Si un estudiante es del sector Urbano, no ejerce actividad laboral y aprueba la asignatura de Compiladores y Sistemas Expertos, es muy probable que también apruebe Control Automatizado Asistido por Computadores.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar y de formar parte del sector Urbano, recalcando así que hay muy pocos estudiantes del sector Rural, están muy presentes dentro de las premisas. De la misma manera, se indica una estrecha relación de aprobación entre asignaturas, para su aprobación en paralelo, no obstante, no se muestra un género femenino pese que se ha considerado una confianza superior al 90%.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas son muy similares a los resultados obtenidos en la herramienta weka. Se mantiene la relación entre la genero, sector de procedencia, el tipo de colegio, si trabaja o no, y si este aprueba materias vinculadas a otras. Esto se ve reflejado en la Figura 5.

## AssociationRules

### Association Rules

```
[trabaja] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.846)
[trabaja] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.846)
[COMPILADORES, trabaja] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.846)
[SISTEMAS EXPERTOS, trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.880)
[SISTEMAS EXPERTOS, trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.880)
[SISTEMAS EXPERTOS, COMPILADORES, trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.880)
[trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.885)
[trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.885)
[COMPILADORES, trabaja] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.885)
[genero, tipo_colegio] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.892)
[provincia_nacimiento] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.900)
[SISTEMAS EXPERTOS, genero, tipo_colegio] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.906)
[genero, tipo_colegio] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.908)
[genero] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.910)
[genero] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.920)
[CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, genero, tipo_colegio] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.921)
[genero, tipo_colegio] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.923)
[genero] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.925)
[provincia_nacimiento] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES] (confidence: 0.925)
[provincia_nacimiento] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.925)
[provincia_nacimiento] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.925)
[edad, genero] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.928)
[edad, genero] --> [SISTEMAS EXPERTOS, CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.928)
[CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, edad, genero] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.928)
[SISTEMAS EXPERTOS, genero] --> [CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, COMPILADORES] (confidence: 0.928)
[CONTROL AUTOMATIZADO ASISTIDO POR COMPUTADORES, genero] --> [SISTEMAS EXPERTOS, COMPILADORES] (confidence: 0.933)
```

Figura 227. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se ha demostrado que para aprobar a las materias relacionadas con las matemáticas deben estar vinculadas con las variables tales como: si el estudiante está o no ejerciendo alguna actividad laboral, la cual se identifica en la mayor parte de las reglas obtenidas que, para que estas sean aprobadas el estudiante debe dedicarse únicamente a las actividades académicas que le corresponden, por otra parte, aprobará a las mismas si el estudiante pertenece al sector Urbano y es oriundo de la provincia de Loja y finalmente se da con la aprobación en paralelo en conjunto con otras asignaturas, esto en relación a características mencionadas con anterioridad.

```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardaro...
+ Código + Texto
RelationRecord(items=frozenset({'C_A', 'CAAC_A', 'Urbano', '11'}), support=0.7326530612244898, ordered_statistics=[Or
RelationRecord(items=frozenset({'NTrabaja', 'CAAC_A', '11', 'SE_A'}), support=0.753061224489796, ordered_statistics=[
RelationRecord(items=frozenset({'C_A', 'NTrabaja', '11', 'SE_A'}), support=0.7306122448979592, ordered_statistics=[Or
RelationRecord(items=frozenset({'C_A', 'SE_A', 'Urbano', '11'}), support=0.7408163265306122, ordered_statistics=[Orde
RelationRecord(items=frozenset({'C_A', 'NTrabaja', 'CAAC_A', 'SE_A'}), support=0.8775510204081632, ordered_statistics:
RelationRecord(items=frozenset({'C_A', 'NTrabaja', 'Urbano', 'CAAC_A'}), support=0.8469387755102041, ordered_statisti
RelationRecord(items=frozenset({'C_A', 'SE_A', 'Urbano', 'CAAC_A'}), support=0.8836734693877552, ordered_statistics=[
RelationRecord(items=frozenset({'NTrabaja', 'SE_A', 'Urbano', 'CAAC_A'}), support=0.8775510204081632, ordered_statist
RelationRecord(items=frozenset({'NTrabaja', 'CAAC_A', 'SE_A', '11', 'C_A'}), support=0.7204081632653061, ordered_stat

```

Figura 228. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances	
0	208 ( 42%)
1	120 ( 24%)
2	110 ( 22%)
3	22 ( 4%)
4	30 ( 6%)

Figura 229. Clúster con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0 y 1. La Figura 8, indica las características que componen a cada clúster, a continuación, se describe cada uno de ellos:

Final cluster centroids:

Attribute	Full Data (490.0)	Cluster#				
		0 (208.0)	1 (120.0)	2 (110.0)	3 (22.0)	4 (30.0)
edad	34	29	31	32	30	34
genero	0	0	0	1	0	1
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	0	0	0
tipo_colegio	1	1	2	1	2	1
COMPILADORES	1	1	1	1	1	1
CONTROL_AUTOMATIZADO_ASISTIDO_POR_COMPUTADORES	1	1	1	1	1	1
SISTEMAS_EXPERTOS	1	1	1	1	1	1

Figura 230. Clústeres obtenidos por Weka.



**Clúster\_0 (42%):** Estudiantes que tengan una edad de 29 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y, que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al décimo ciclo, es decir, Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos.

**Clúster\_1 (24%):** Estudiantes que tengan una edad de 31 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y, que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al décimo ciclo, es decir, Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos.

**Clúster\_2 (22%):** Estudiantes que tengan una edad de 32 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y, que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al décimo ciclo, es decir, Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos.

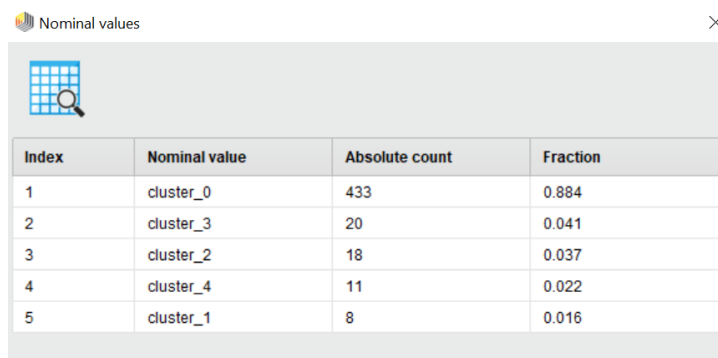
**Clúster\_3 (4%):** Estudiantes que tengan una edad de 30 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja, sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y, que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al décimo ciclo, es decir, Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos.

**Clúster\_4 (6%):** Estudiantes que tengan una edad de 34 años, pertenezcan al género masculino, de estado civil soltero, provenientes de la provincia de Loja,

sean del sector Urbano, que no cuentan con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y, que hayan finalizado su secundaria en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al décimo ciclo, es decir, Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

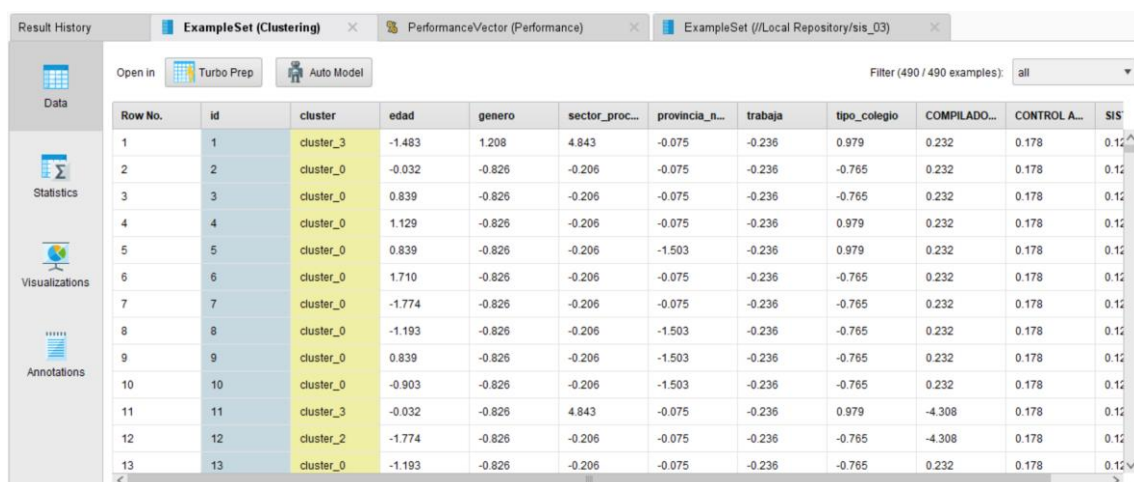
La Figura 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	433	0.884
2	cluster_3	20	0.041
3	cluster_2	18	0.037
4	cluster_4	11	0.022
5	cluster_1	8	0.016

Figura 231. Clústers con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 314 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	COMPILADO...	CONTROL A...	SIS
1	1	cluster_3	-1.483	1.208	4.843	-0.075	-0.236	0.979	0.232	0.178	0.12
2	2	cluster_0	-0.032	-0.826	-0.206	-0.075	-0.236	-0.765	0.232	0.178	0.12
3	3	cluster_0	0.839	-0.826	-0.206	-0.075	-0.236	-0.765	0.232	0.178	0.12
4	4	cluster_0	1.129	-0.826	-0.206	-0.075	-0.236	0.979	0.232	0.178	0.12
5	5	cluster_0	0.839	-0.826	-0.206	-1.503	-0.236	0.979	0.232	0.178	0.12
6	6	cluster_0	1.710	-0.826	-0.206	-0.075	-0.236	-0.765	0.232	0.178	0.12
7	7	cluster_0	-1.774	-0.826	-0.206	-0.075	-0.236	-0.765	0.232	0.178	0.12
8	8	cluster_0	-1.193	-0.826	-0.206	-1.503	-0.236	-0.765	0.232	0.178	0.12
9	9	cluster_0	0.839	-0.826	-0.206	-1.503	-0.236	-0.765	0.232	0.178	0.12
10	10	cluster_0	-0.903	-0.826	-0.206	-1.503	-0.236	-0.765	0.232	0.178	0.12
11	11	cluster_3	-0.032	-0.826	4.843	-0.075	-0.236	0.979	-4.308	0.178	0.12
12	12	cluster_2	-1.774	-0.826	-0.206	-0.075	-0.236	-0.765	-4.308	0.178	0.12
13	13	cluster_0	-1.193	-0.826	-0.206	-0.075	-0.236	-0.765	0.232	0.178	0.12

Figura 232. Clúster con total de instancias



De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 1 con más instancias agrupadas.

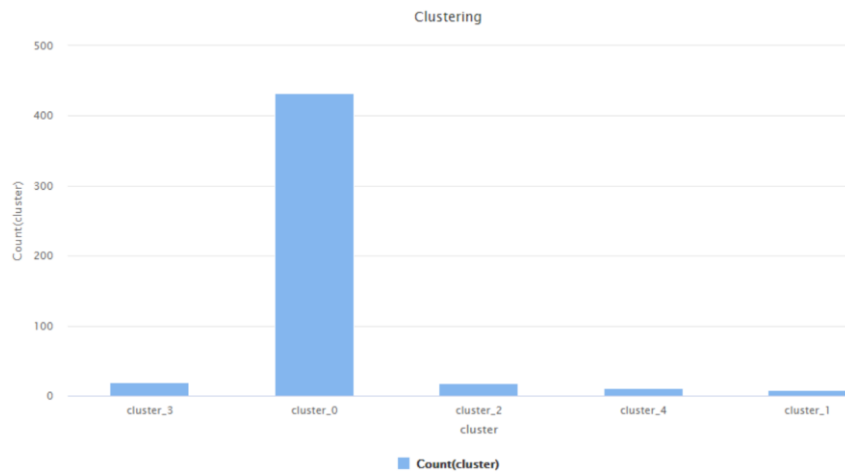


Figura 233. Representación gráfica de los clústers obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una con sus características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	COMPILADO...	CONTROL A...	edad	genero	provincia_n...	sector_proc...	SISTEMAS E...	tipo_colegio	trabaja
Cluster 0	0.942	1	31.413	1	11.407	0	1	1.344	0.016
Cluster 1	0.500	0.500	30.000	0.500	13.500	0	0.000	1.250	0.125
Cluster 2	0.950	1	28.200	0.150	11.400	1	1	1.450	0.050
Cluster 3	0.818	0.000	29.091	0.273	11.364	0	1	1.545	0.273
Cluster 4	0.973	1	31.233	0	10.977	0	1	1.508	0.069

Figura 234. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos, todas pertenecientes al décimo ciclo, si tiene una edad de 31 años, es de género femenino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 1:** Estudiante reprueba todas las asignaturas relacionadas a las matemáticas como Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos, todas pertenecientes al décimo ciclo, si tiene una edad de 30 años, es de género masculino, oriundo de la provincia de Manabí, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 2:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos, todas pertenecientes al décimo ciclo, si tiene una edad de 28 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Rural, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 3:** Estudiante aprueba dos asignaturas relacionadas a las matemáticas como Compiladores, y Sistemas Expertos, y reprueba la materia de Control Automatizado Asistido por Computadores todas pertenecientes al décimo ciclo, si tiene una edad de 29 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscal, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

**Clúster 4:** Estudiante aprueba todas las asignaturas relacionadas a las matemáticas como Compiladores, Control Automatizado Asistido por Computadores y Sistemas Expertos, todas pertenecientes al décimo ciclo, si tiene una edad de 31 años, es de género masculino, oriundo de la provincia de Loja, perteneciente al sector Urbano, proviene de un colegio Fiscomisional, y únicamente se dedica a sus actividades académicas, es decir, no trabaja.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.





Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Análisis numérico**, donde se observa que un grupo de estudiantes pertenecientes al género femenino aprueban la materia, por otro lado, el género masculino también aprueba a dicha asignatura, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprobaban la materia de Control Automatizado Asistido por computadores.

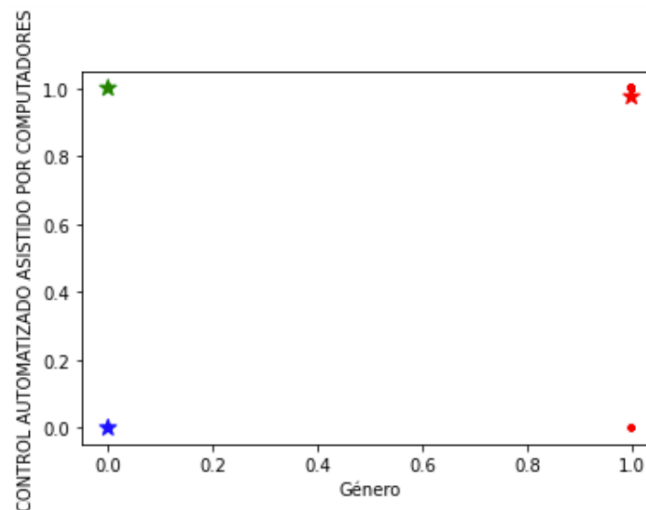


Figura 237. Relación Género – Control Automatizado Asistido por computadores

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Compiladores** y **Control Automatizado Asistido por Computadores**, en el clúster verde se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde apruebe únicamente a Control Automatizado Asistido por Computadores (clúster azul), y por último (clúster verde) se muestra la probabilidad de que apruebe la materia de Compiladores pero reprobé la materia antes mencionada.

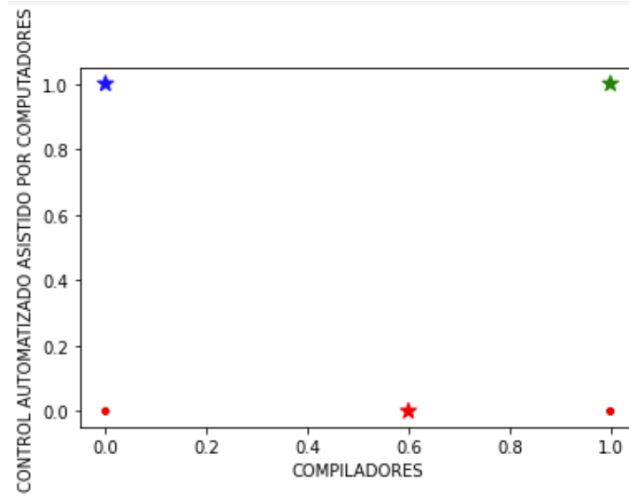


Figura 238. Relación Compiladores – Control Automatizado Asistido por computadores

## Anexo 14. Ingeniería en Computación – Primer ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **primer ciclo de la Carrera de Ingeniería en Computación** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

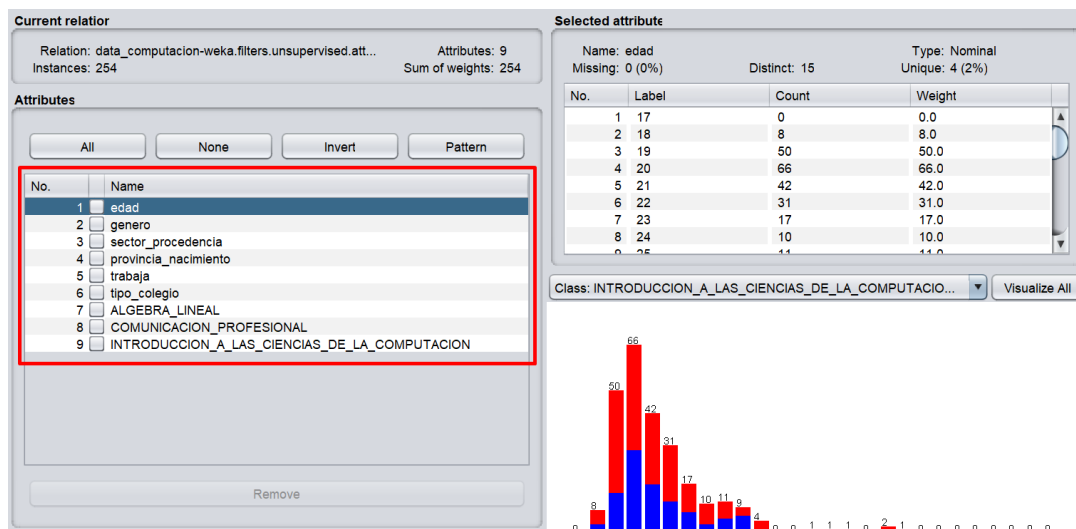


Figura 239. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. genero=0 tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 116 ==> trabaja=0 COMUNICACION_PROF
177. genero=0 192 ==> sector_procedencia=0 177 <conf:(0.92)> lift:(1.03) lev:(0.02) [5] conv:(1.28)
178. genero=0 192 ==> trabaja=0 tipo_colegio=2 177 <conf:(0.92)> lift:(1) lev:(-0) [0] conv:(0.9)
179. genero=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 115 ==> tipo_colegio=2 106 <conf:(0.92)> lift:(0.97)
180. genero=0 provincia_nacimiento=11 COMUNICACION_PROFESIONAL=1 114 ==> sector_procedencia=0 105 <conf:(0.92)
181. provincia_nacimiento=11 tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 114 ==> COMUNICACION
182. genero=0 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 126 ==> tipo_colegio=2 116 <conf:(0.92)> lift:(
183. provincia_nacimiento=11 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 124 ==> tipo_colegio=2 114 <conf
184. provincia_nacimiento=11 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 111 ==> tipo_colegio=2 102 <conf:(0.
185. genero=0 tipo_colegio=2 COMUNICACION_PROFESIONAL=1 135 ==> sector_procedencia=0 124 <conf:(0.92)> lift:(
186. provincia_nacimiento=11 208 ==> trabaja=0 tipo_colegio=2 191 <conf:(0.92)> lift:(0.99) lev:(-0.01) [-1]
187. genero=0 COMUNICACION_PROFESIONAL=1 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 121 ==> tipo_colegio=2
188. genero=0 provincia_nacimiento=11 tipo_colegio=2 145 ==> sector_procedencia=0 trabaja=0 133 <conf:(0.92)> lift:(
189. genero=0 provincia_nacimiento=11 155 ==> trabaja=0 tipo_colegio=2 142 <conf:(0.92)> lift:(0.99) lev:(-0.
190. genero=0 trabaja=0 tipo_colegio=2 COMUNICACION_PROFESIONAL=1 131 ==> sector_procedencia=0 120 <conf:(0.9
191. sector_procedencia=0 COMUNICACION_PROFESIONAL=1 164 ==> trabaja=0 tipo_colegio=2 150 <conf:(0.91)> lift:(
192. genero=0 tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 116 ==> sector_procedencia=0 106
193. genero=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 115 ==> INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=
194. provincia_nacimiento=11 COMUNICACION_PROFESIONAL=1 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 115 ==>
195. genero=0 trabaja=0 COMUNICACION_PROFESIONAL=1 137 ==> sector_procedencia=0 125 <conf:(0.91)> lift:(1.02)
196. genero=0 provincia_nacimiento=11 COMUNICACION_PROFESIONAL=1 114 ==> tipo_colegio=2 104 <conf:(0.91)> lif
197. ALGEBRA_LINEAL=1 158 ==> trabaja=0 tipo_colegio=2 144 <conf:(0.91)> lift:(0.99) lev:(-0.01) [-2] conv:(0
198. provincia_nacimiento=11 tipo_colegio=2 COMUNICACION_PROFESIONAL=1 135 ==> sector_procedencia=0 123 <conf
199. genero=0 trabaja=0 tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 112 ==> sector_procedenci
200. genero=0 COMUNICACION_PROFESIONAL=1 145 ==> sector_procedencia=0 132 <conf:(0.91)> lift:(1.02) lev:(0.01
  
```

Figura 240. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados. Aplicado los filtros para reglas con nivel de confianza mayor a 97% y lift mayor a 1 respectivamente, se obtiene un total de 20 reglas (Ver Figura 3).

REGLAS	CONF	LIFT			
1. genero=0 provincia_nacimiento=11 tipo_colegio=2 145 ==> trabaja=0 142	conf:(0.98)>	lift:(1.02)	lev:(0.01)	[3]	conv:(1.57)
2. genero=0 sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=2 136 ==> trabaja=0 133	conf:(0.98)>	lift:(1.02)	lev:(0.01)	[2]	conv:(1.47)
3. genero=0 tipo_colegio=2 181 ==> trabaja=0 177	conf:(0.98)>	lift:(1.02)	lev:(0.02)	[3]	conv:(1.57)
4. genero=0 sector_procedencia=0 tipo_colegio=2 168 ==> trabaja=0 164	conf:(0.98)>	lift:(1.02)	lev:(0.01)	[3]	conv:(1.46)
5. sector_procedencia=0 trabaja=0 217 ==> tipo_colegio=2 211	conf:(0.97)>	lift:(1.02)	lev:(0.02)	[4]	conv:(1.46)
6. sector_procedencia=0 provincia_nacimiento=11 trabaja=0 177 ==> tipo_colegio=2 172	conf:(0.97)>	lift:(1.02)	lev:(0.01)	[3]	conv:(1.39)
7. tipo_colegio=2 242 ==> trabaja=0 235	conf:(0.97)>	lift:(1.02)	lev:(0.01)	[3]	conv:(1.31)
8. genero=0 tipo_colegio=2 COMUNICACION_PROFESIONAL=1 135 ==> trabaja=0 131	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.17)
9. provincia_nacimiento=11 tipo_colegio=2 197 ==> trabaja=0 191	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[2]	conv:(1.22)
10. sector_procedencia=0 tipo_colegio=2 218 ==> trabaja=0 211	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[2]	conv:(1.18)
11. sector_procedencia=0 trabaja=0 COMUNICACION_PROFESIONAL=1 155 ==> tipo_colegio=2 150	conf:(0.97)>	lift:(1.02)	lev:(0.01)	[2]	conv:(1.22)
12. genero=0 sector_procedencia=0 tipo_colegio=2 COMUNICACION_PROFESIONAL=1 124 ==> trabaja=0 121	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.07)
13. trabaja=0 243 ==> tipo_colegio=2 235	conf:(0.97)>	lift:(1.02)	lev:(0.01)	[3]	conv:(1.28)
14. tipo_colegio=2 ALGEBRA_LINEAL=1 149 ==> trabaja=0 144	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.08)
15. provincia_nacimiento=11 tipo_colegio=2 ALGEBRA_LINEAL=1 119 ==> trabaja=0 115	conf:(0.97)>	lift:(1.01)	lev:(0)	[1]	conv:(1.03)
16. sector_procedencia=0 provincia_nacimiento=11 tipo_colegio=2 178 ==> trabaja=0 172	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.1)
17. genero=0 tipo_colegio=2 ALGEBRA_LINEAL=1 118 ==> trabaja=0 114	conf:(0.97)>	lift:(1.01)	lev:(0)	[1]	conv:(1.02)
18. tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 147 ==> trabaja=0 142	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.06)
19. genero=0 tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 116 ==> trabaja=0 111	conf:(0.97)>	lift:(1.01)	lev:(0)	[1]	conv:(1)
20. tipo_colegio=2 COMUNICACION_PROFESIONAL=1 172 ==> trabaja=0 166	conf:(0.97)>	lift:(1.01)	lev:(0.01)	[1]	conv:(1.06)

Figura 241. Reglas obtenidas aplicando filtros (Conf – Lift)



A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante de género masculino, pertenece a la provincia de Loja y culmina sus estudios secundarios en una Institución Fiscomisional, es muy probable que tampoco ejerza alguna actividad laboral.
- 2) Si un estudiante de género masculino, es procedente del sector Urbano, pertenece a la provincia de Loja y culmina sus estudios secundarios en una Institución Fiscomisional, es muy probable que tampoco ejerza alguna actividad laboral.
- 3) Estudiante de género masculino y finaliza sus estudios secundarios en una Institución Fiscomisional, es muy probable que tampoco ejerza alguna actividad laboral.
- 4) Si un estudiante de género masculino, procedente del sector Urbano y culmina sus estudios secundarios en una Institución Fiscomisional, es muy probable que tampoco ejerza alguna actividad laboral.
- 5) Estudiante procedente del sector Urbano, no ejerce actividad laboral, es muy probable que tampoco ejerza alguna actividad laboral.
- 6) Estudiante procedente del sector Urbano, oriundo de la provincia de Loja y no ejerce actividad laboral, es muy probable que provenga de una Institución Fiscomisional.
- 7) Estudiante que haya culminado sus estudios secundarios en una Institución Fiscomisional, es muy probable que la actualidad no ejerza ninguna actividad laboral.
- 8) Estudiante de género masculino, que haya culminado sus estudios secundarios en una Institución Fiscomisional y apruebe la materia de Comunicación Profesional, es muy probable que no cuente con ninguna actividad laboral.
- 9) Si un estudiante oriundo de la provincia de Loja y culmina sus estudios en una Institución Fiscomisional, es muy probable que no ejerza alguna actividad laboral.
- 10) Estudiante que es procedente del sector Urbano y culmina sus estudios secundarios en una Institución Fiscomisional, es muy probable que tampoco ejerza alguna actividad laboral.



- 11) Si un estudiante es procedente del sector Urbano, no ejerce actividad laboral, aprueba la Asignatura de Comunicación Profesional, es muy probable que sus estudios secundarios los haya finalizado en una institución Fiscomisional.
- 12) Estudiante de género masculino, que pertenezca al sector Urbano, que haya culminado sus estudios secundarios en una Institución Fiscomisional y que apruebe la materia de Comunicación Profesional, es muy probable que no ejerza alguna actividad laboral.
- 13) Si un estudiante no cuenta con ninguna actividad laboral, es muy probable que sus estudios secundarios los haya culminado en una Institución Fiscomisional.
- 14) Si un estudiante que haya terminado sus estudios secundarios en una Institución Fiscomisional y aprueba la asignatura de Algebra Lineal, es muy probable que no ejerza con ninguna actividad laboral.
- 15) Si un estudiante es oriundo de la provincia de Loja y culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Algebra Lineal, es muy probable que no realice ninguna actividad laboral.
- 16) Si un estudiante pertenece al sector Urbano, oriundo de la provincia de Loja y culmina sus estudios en una Institución Fiscomisional es muy probable que no realice ninguna actividad laboral.
- 17) Estudiante de género masculino, ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Algebra Lineal, es muy probable que no realice ninguna actividad laboral.
- 18) Si un estudiante culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la asignatura de Introducción a las Ciencias de la Computación, es muy probable que no realice ninguna actividad laboral.
- 19) Estudiante de género masculino, culminado sus estudios en una Institución Fiscomisional y aprueba la materia de Introducción a las Ciencias de la Computación, es muy probable que no realice ninguna actividad laboral.
- 20) Si un estudiante culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Comunicación Profesional, es muy probable que no realice ninguna actividad laboral.



Todas las reglas obtenidas considerando a la confidence (confianza) como métrica, muestran una alta confiabilidad entre las mismas, además se recalca la relación con el lift (elevación), el cual indica cuándo una regla es mejor prediciendo el resultado que asumiendo dicho resultado de forma aleatoria, por ello se identifica que ambas medidas se encuentran correlacionadas siendo en este caso la confianza con más peso.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado, obteniendo valores mayores a 1.45 y una confiabilidad de las premisas desde el 67% al 91% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONF	LIFT		
1. ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 140 ==> genero=0 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 105 -(0.75)		lift:(1.51)>	lev:(0.14)	[35]
2. genero=0 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 126 ==> ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 105 -(0.83)		lift:(1.51)>	lev:(0.14)	[35]
3. INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 157 ==> genero=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 105 -(0.67)		lift:(1.48)>	lev:(0.13)	[33]
4. genero=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 115 ==> INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 105 -(0.91)		lift:(1.48)>	lev:(0.13)	[33]
5. trabaja=0 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 148 ==> tipo_colegio=2 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 105 -(0.75)		lift:(1.45)>	lev:(0.14)	[34]
6. tipo_colegio=2 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 131 ==> trabaja=0 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 105 -(0.85)		lift:(1.45)>	lev:(0.14)	[34]
7. trabaja=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 132 ==> tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 105 -(0.84)		lift:(1.45)>	lev:(0.14)	[34]
8. tipo_colegio=2 INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 147 ==> trabaja=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 105 -(0.76)		lift:(1.45)>	lev:(0.14)	[34]
9. sector_procedencia=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 125 ==> INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 105 -(0.9)		lift:(1.45)>	lev:(0.14)	[34]
10. INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION=1 157 ==> sector_procedencia=0 ALGEBRA_LINEAL=1 COMUNICACION_PROFESIONAL=1 105 -(0.71)		lift:(1.45)>	lev:(0.14)	[34]

Figura 242. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante aprueba a las asignaturas: Álgebra Lineal y Comunicación Social, es muy probable que sea de género masculino y que apruebe también a la asignatura de Introducción a las Ciencias de la Computación.
- 2) Estudiante de género masculino y que apruebe la asignatura de Introducción a las Ciencias de la Computación, es muy probable que también apruebe a las asignaturas de Álgebra Lineal y Comunicación Social.
- 3) Si un estudiante aprueba la materia de Introducción a las Ciencias de la Computación, es muy probable que sea de género masculino y que también apruebe a las materias de Álgebra Lineal y Comunicación Social.
- 4) Estudiante de género masculino y que apruebe la asignatura de Álgebra Lineal y Comunicación Social, existe la posibilidad de que también apruebe la materia de Introducción a las Ciencias de la Computación.
- 5) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Introducción a las Ciencias de la Computación, es muy probable que haya culminado sus estudios secundarios en una Institución Fiscomisional y que apruebe a las materias de Álgebra Lineal y Comunicación Social.



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

- 6) Estudiante que haya culminado sus estudios secundarios en Institución Fiscomisional y aprueba las materias de Algebra Lineal y Comunicación Social, es muy probable que no realice ninguna actividad laboral y también apruebe a la materia de Introducción a las Ciencias de la Computación.
- 7) Estudiante que no realice ninguna actividad laboral y aprueba las materias de Algebra Lineal y Comunicación Social, es muy probable que haya culminado sus estudios secundarios en Institución Fiscomisional y también apruebe a la materia de Introducción a las Ciencias de la Computación.
- 8) Estudiante que haya culminado sus estudios secundarios en Institución Fiscomisional y aprueba la materia de Introducción a las Ciencias de la Computación, es muy probable que no realice ninguna actividad laboral y también apruebe a las asignaturas de Algebra Lineal y Comunicación Social.
- 9) Estudiante que pertenezca al sector Urbano y aprueba las materias de Algebra Lineal y Comunicación Social, es muy probable que no realice ninguna actividad laboral y también apruebe a la materia de Introducción a las Ciencias de la Computación.
- 10) Si un estudiante aprueba la materia de Introducción a las Ciencias de la Computación, es muy probable que pertenezca al sector Urbano y que también apruebe a las materias de Algebra Lineal y Comunicación Social.

En este caso al tomar al lift como métrica principal, se obtiene reglas con resultados mejor predictivos que al azar o de manera random. Sin embargo, se identifica por ejemplo a las reglas # 3 y 10, que los estudiantes que aprueban a la asignatura de Introducción a las Ciencias de la Computación, deben ser de sexo masculino, pertenecer al sector Urbano y aprobar a las materias de Algebra Lineal y Comunicación Social, estas cuentan con un lift alto de 1.45 y 1.48 pero con una confianza de 67% y 71%, si bien es cierto las reglas están correlacionadas, pero la tasa de confiabilidad no es la mejor en comparación a las reglas restantes obtenidas, por ejemplo el caso de las reglas # 4 y 9, indican una relación contraria en las reglas expuestas anteriormente, en este caso indica que el estudiante debe ser de género masculino y aprobar a las materias de Algebra Lineal y Comunicación Social para que sea aprobada la materia de





Introducción a las Ciencias de la computación, la relación va la siguiente regla va estructurada de la misma manera pero a esta se le añade de que dicho estudiante debe pertenecer al sector Urbano, esto refleja que ambas medidas están fuertemente correlacionadas y las premisas se efectuarán de una manera más precisa y confiable.

Además, cada una de las reglas obtenidas también muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de pertenecer a un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. Sin embargo, la relación de las premisas con el género femenino en particular es muy bajo.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en weka. La relación entre las asignaturas se sigue manteniendo, se muestra la presencia de estudiante provenientes de la Provincia de Loja, además, en este caso se muestra la característica tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 32 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

## AssociationRules

```
Association Rules
[COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, provincia_nacimiento, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [edad, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
```

Figura 243. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se logra identificar que la relación entre las asignaturas con las variables demográficas o sociales es muy baja, ya que el nivel de confiabilidad en las mismas no alcanza el 70 %, mientras tanto que, la relación únicamente en variables sociales y demográficas es muy alta, estas premisas tienen una confianza de hasta el 91%, con una predicción para el futuro de 1.014. Las características que se identifican con más periodicidad dentro de las premisas son aquellos estudiantes de género masculino, provenientes del sector Urbano y de un colegio Privado, son solteros y no cuentan con ninguna actividad laboral, estas son las características más relevantes que se destacan dentro de las premisas.

```

Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardaron todos los cambios
Comentar Compartir Editando
+ Código + Texto
RelationRecord(items=frozenset({'privado', 'CP_A', 'Urbano'}), support=0.6141732283464567, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Urbano'}),
RelationRecord(items=frozenset({'NTrabaja', 'Masculino', 'Urbano'}), support=0.6692913385826772, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Masculino', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'Masculino'}), support=0.6968503937907874, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', 'Masculino'}),
RelationRecord(items=frozenset({'Masculino', 'Urbano', 'Soltero'}), support=0.6929133858267716, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Masculino', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'Masculino', 'Urbano'}), support=0.6614173228346457, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Masculino', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'Soltero'}), support=0.9133858267716536, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', 'Soltero'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'Urbano'}), support=0.8307086614173228, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'Urbano', 'Soltero'}), support=0.8503937907874016, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Urbano', 'Soltero'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', '11', 'Soltero'}), support=0.7440944881889764, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', '11', 'Soltero'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', '11', 'Urbano'}), support=0.6771653543307087, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', '11', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'CP_A', 'Urbano', 'Soltero'}), support=0.610236220472441, ordered_statistics=[OrderedStatistic(items_base=frozenset({'CP_A', 'Urbano', 'Soltero'}),
RelationRecord(items=frozenset({'NTrabaja', 'Masculino', 'Urbano', 'Soltero'}), support=0.6653543307086615, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja', 'Masculino', 'Urbano'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'Masculino', 'Soltero'}), support=0.6889763779527559, ordered_statistics=[OrderedStatistic(items_base=frozenset({'privado', 'NTrabaja', 'Masculino'}),
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'Masculino', 'Urbano'}), support=0.6456692913385826, ordered_statistics=[OrderedStatistic(items_base=frozenset({'privado', 'NTrabaja', 'Masculino'}),
RelationRecord(items=frozenset({'privado', 'Masculino', 'Urbano', 'Soltero'}), support=0.65748031496063, ordered_statistics=[OrderedStatistic(items_base=frozenset({'privado', 'Masculino', 'Urbano'}),

```

Figura 244. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

Clustered Instances	
0	86 ( 34%)
1	22 ( 9%)
2	55 ( 22%)
3	84 ( 33%)
4	7 ( 3%)

Figura 245. Clúster con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 2, y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Attribute	Full Data (254.0)	Cluster#				
		0 (86.0)	1 (22.0)	2 (55.0)	3 (84.0)	4 (7.0)
edad	20	20	20	19	20	23
genero	0	0	1	1	0	0
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	0	0	0
tipo_colegio	2	2	2	2	2	2
ALGEBRA_LINEAL	1	0	0	1	1	0
COMUNICACION_PROFESIONAL	1	1	0	1	1	1
INTRODUCCION_A_LAS_CIENCIAS_DE_LA_COMPUTACION	1	0	0	1	1	0

Figura 246. Clústeres obtenidos por Weka.



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

**Clúster\_0 (34%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobará únicamente la asignatura de Comunicación Profesional mientras que, reprobará las materias de Álgebra Lineal e Introducción a las Ciencias de la Computación.

**Clúster\_1 (9%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, reprobará todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, estas son: Álgebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

**Clúster\_2 (22%):** Estudiantes que tengan una edad de 19 años, pertenezcan al género femenino, procedentes del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, estas son: Álgebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

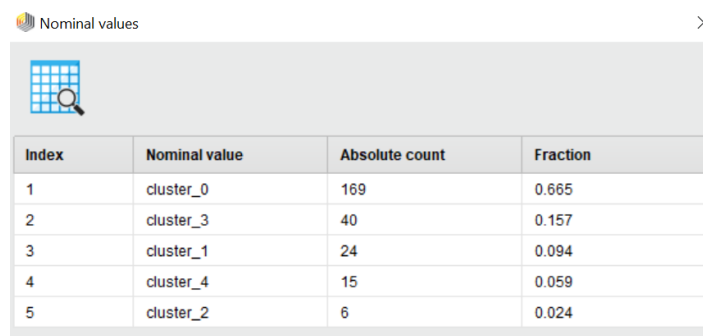
**Clúster\_3 (33%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género masculino, procedentes del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, estas son: Álgebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

**Clúster\_4 (3%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y únicamente se hacen cargo de sus actividades académicas y que hayan finalizado su secundaria en una institución

Fiscomisional, reprobará todas las materias relacionadas a las matemáticas correspondientes al primer ciclo, estas son: Algebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

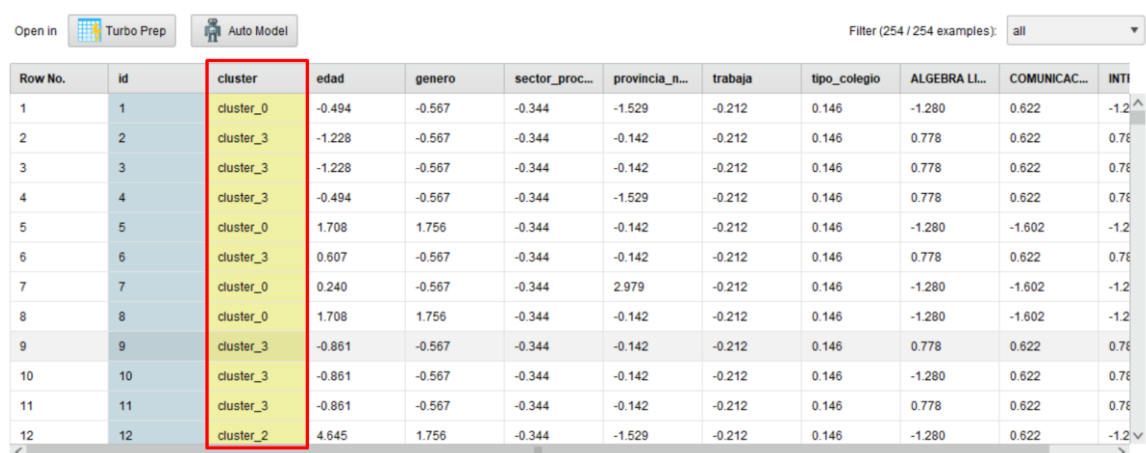
La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústeres.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	169	0.665
2	cluster_3	40	0.157
3	cluster_1	24	0.094
4	cluster_4	15	0.059
5	cluster_2	6	0.024

Figura 247. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 254 instancias están clasificadas o agrupadas a su clúster más cercano.



Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	ALGEBRA LI...	COMUNICAC...	INTI
1	1	cluster_0	-0.494	-0.567	-0.344	-1.529	-0.212	0.146	-1.280	0.622	-1.2
2	2	cluster_3	-1.228	-0.567	-0.344	-0.142	-0.212	0.146	0.778	0.622	0.76
3	3	cluster_3	-1.228	-0.567	-0.344	-0.142	-0.212	0.146	0.778	0.622	0.76
4	4	cluster_3	-0.494	-0.567	-0.344	-1.529	-0.212	0.146	0.778	0.622	0.76
5	5	cluster_0	1.708	1.756	-0.344	-0.142	-0.212	0.146	-1.280	-1.602	-1.2
6	6	cluster_3	0.607	-0.567	-0.344	-0.142	-0.212	0.146	0.778	0.622	0.76
7	7	cluster_0	0.240	-0.567	-0.344	2.979	-0.212	0.146	-1.280	-1.602	-1.2
8	8	cluster_0	1.708	1.756	-0.344	-0.142	-0.212	0.146	-1.280	-1.602	-1.2
9	9	cluster_3	-0.861	-0.567	-0.344	-0.142	-0.212	0.146	0.778	0.622	0.76
10	10	cluster_3	-0.861	-0.567	-0.344	-0.142	-0.212	0.146	-1.280	0.622	0.76
11	11	cluster_3	-0.861	-0.567	-0.344	-0.142	-0.212	0.146	0.778	0.622	0.76
12	12	cluster_2	4.645	1.756	-0.344	-1.529	-0.212	0.146	-1.280	0.622	-1.2

Figura 248. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.

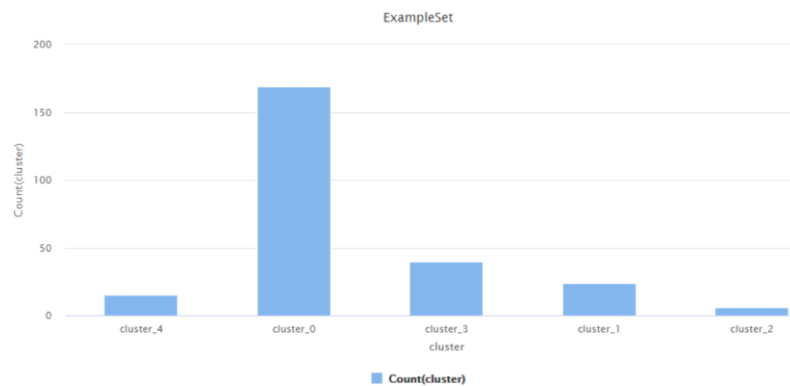


Figura 249. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	ALGEBRA LI...	COMUNICAC...	edad	genero	INTRODUCCI...	provincia_n...	sector_proc...	tipo_colegio	trabaja
Cluster 0	0.125	0.238	21.637	0.300	0.025	11.313	0.100	2.000	-0
Cluster 1	0.667	1	29.667	0	1	11.000	0.333	1	1
Cluster 2	0.750	0.875	23.500	0.375	0.750	10.500	0	2.125	1
Cluster 3	0.865	0.949	20.872	0.218	0.904	11.545	0.103	2.006	0.000
Cluster 4	0.714	0.857	22.571	0.143	0.714	10.714	0.286	1	0

Figura 250. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante masculino de 21 años de edad, oriundo de la Provincia de Loja, no ejerce ninguna actividad laboral, reside en el sector Urbano y culmina sus estudios secundarios en una Institución Fiscomisional, reprueba todas las asignaturas relacionadas a las matemáticas como Algebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

**Clúster 1:** Estudiante masculino de 30 años de edad, oriundo de la Provincia de Loja, cuenta con una actividad laboral, reside en el sector Urbano y culmina sus





estudios secundarios en una Institución Fiscal, aprueba todas las dos de las asignaturas relacionadas a las matemáticas como son: Comunicación Profesional e Introducción a las Ciencias de la Computación y tiene una probabilidad de aprobar a la materia de Algebra Lineal.

**Clúster 2:** Estudiante masculino de 23 años de edad, oriundo de la Provincia de Loja, cuenta con una actividad laboral, reside en el sector Urbano y culmina sus estudios secundarios en una Institución Fiscomisional, tiene una alta probabilidad de aprobar todas las asignaturas relacionadas a las matemáticas como Algebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

**Clúster 3:** Estudiante masculino de 20 años de edad, oriundo de la Provincia de Loja, no ejerce ninguna actividad laboral, reside en el sector Urbano y culmina sus estudios secundarios en una Institución Fiscomisional, tiene una probabilidad de aprobar todas las asignaturas relacionadas a las matemáticas como Algebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

**Clúster 4:** Estudiante masculino de 22 años de edad, oriundo de la Provincia de Loja, no ejerce ninguna actividad laboral, reside en el sector Urbano y culmina sus estudios secundarios en una Institución Fiscal, tiene una probabilidad de aprobar todas las asignaturas relacionadas a las matemáticas como Algebra Lineal, Comunicación Profesional e Introducción a las Ciencias de la Computación.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide rojo y verde, se observa una relación entre los estudiantes de género femenino y masculino respectivamente, con una Institución Fiscomisional, mientras que, en el centroide de color azul hace referencia a estudiantes de género Masculino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, sin embargo, en instancias anteriores si se visualiza la presencias de instituciones privadas.

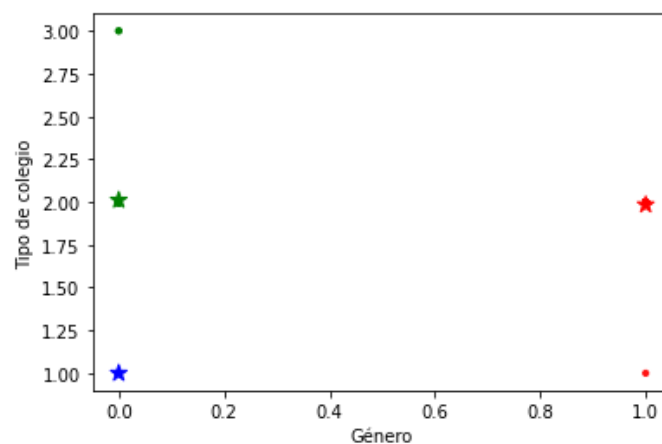


Figura 251. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Algebra Lineal** en donde, los estudiantes aprueban a la asignatura con una edad de 20, 23 y 32 años.

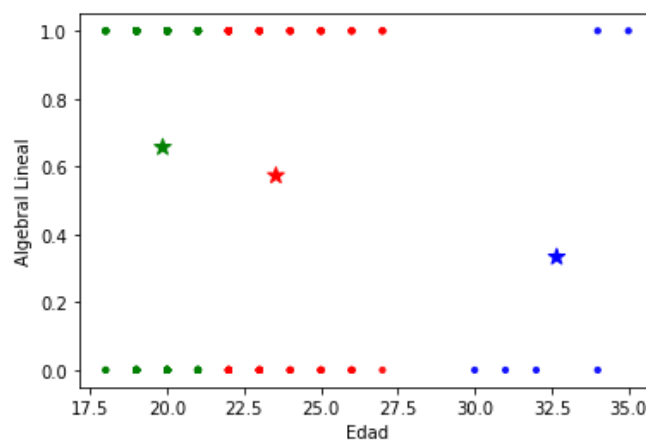


Figura 252. Relación Edad – Ingeniería de Software II



Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género y Algebra Lineal**, donde se observa que un grupo de estudiantes pertenecientes al género femenino (centroide verde) cuenta con una probabilidad (baja) de aprobar a dicha asignatura, por otro lado, el género masculino aprueba la asignatura de Algebra Lineal (centroide rojo), y considerando al lado centroide azul, se muestra todo lo contrario, es decir, reprueban.

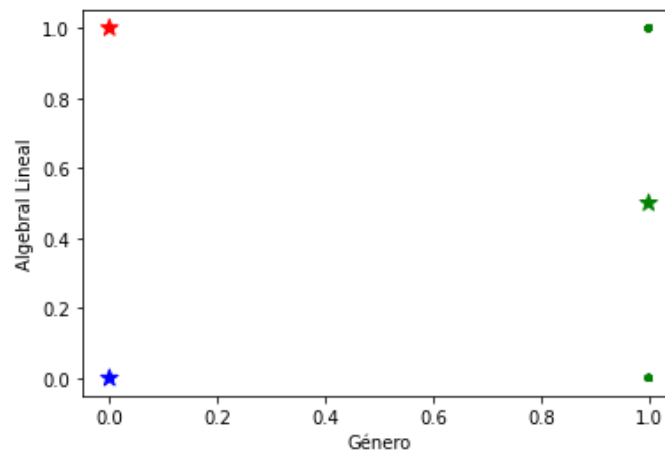


Figura 253. Relación Género – Ingeniería de Software II

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Introducción a las Ciencias de la Computación y Algebra Lineal**, en el clúster azul se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde únicamente apruebe Algebra Lineal y repruebe Introducción a las Ciencias de la Computación (clúster rojo), finalmente se muestra una baja probabilidad que el estudiante apruebe la materia de Introducción a las Ciencias de la Computación (clúster verde).

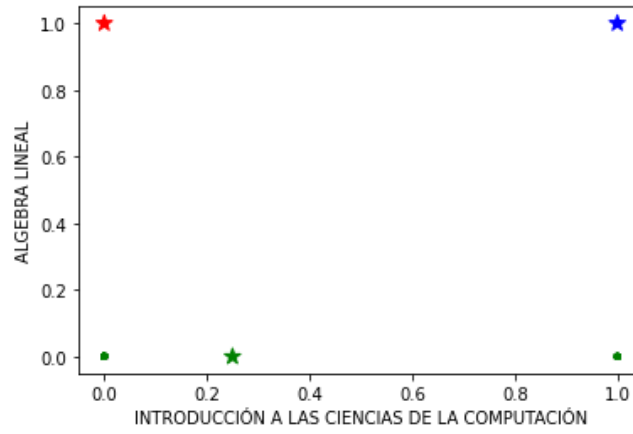


Figura 254. Relación Modelamiento Matemático – Ingeniería de Software II

## Anexo 15. Ingeniería en Computación – Segundo ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **Segundo ciclo de la Carrera de Ingeniería en Computación** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas Weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

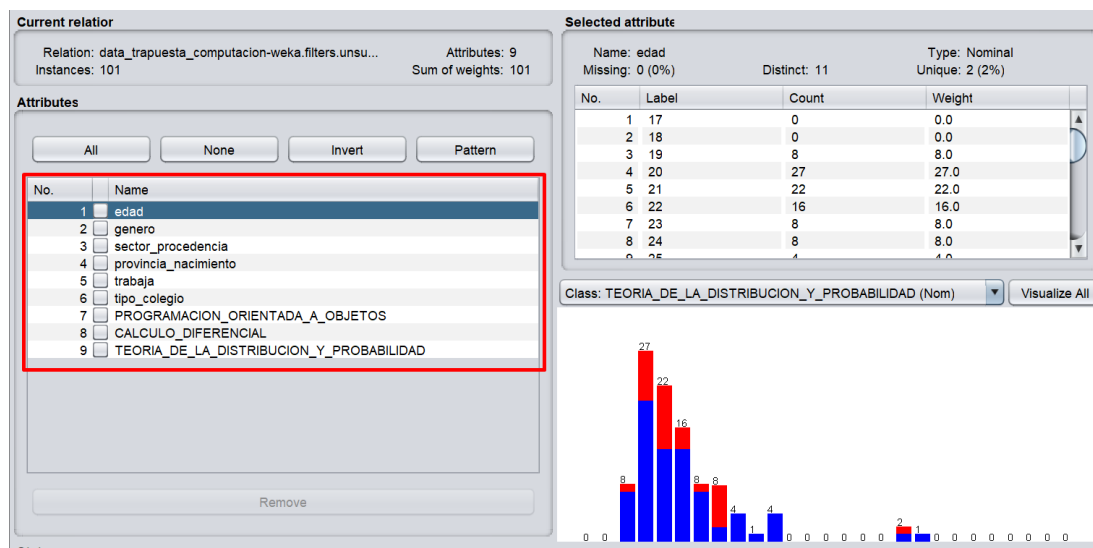


Figura 255. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a

tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```

Associator output
176. genero=0 provincia_nacimiento=11 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 44 ==> sector_procedencia=0 40
177. sector_procedencia=0 provincia_nacimiento=11 trabaja=0 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 44 ==> ti
178. genero=0 trabaja=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 44 ==> s
179. genero=0 sector_procedencia=0 trabaja=0 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 44
180. genero=0 sector_procedencia=0 trabaja=0 tipo_colegio=2 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 44 ==> CA
181. trabaja=0 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 65 ==> sector_procedencia=0 59 <conf:(0.91)> lift:(0
182. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 54 ==> sector_procedencia=0 49 <conf:(0.91)> lift:(1.0
183. genero=0 trabaja=0 CALCULO_DIFERENCIAL=1 54 ==> sector_procedencia=0 49 <conf:(0.91)> lift:(1.03) lev:(0
184. genero=0 trabaja=0 CALCULO_DIFERENCIAL=1 54 ==> tipo_colegio=2 49 <conf:(0.91)> lift:(1) lev:(-0) [0] cc
185. genero=0 trabaja=0 CALCULO_DIFERENCIAL=1 54 ==> TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 49 <conf:(0.9
186. trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 CALCULO_DIFERENCIAL=1 54 ==> tipo_colegio=2 49 <conf:(0.91)
187. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 54 ==> trabaja=0 CALCULO_DIFERENCIAL=1 49 <conf:(0.91)
188. genero=0 trabaja=0 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 53 ==> sector_procedencia=0 48 <conf:(0.91)
189. genero=0 trabaja=0 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 53 ==> tipo_colegio=2 48 <conf:(0.91)> lif
190. provincia_nacimiento=11 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 53 ==> sector_pro
191. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 CALCULO_DIFERENCIAL=1 53 ==> sector_procedencia=0 48 <
192. trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 53 ==> tipo_colegio
193. trabaja=0 tipo_colegio=2 CALCULO_DIFERENCIAL=1 63 ==> sector_procedencia=0 57 <conf:(0.9)> lift:(1.03) l
194. TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 73 ==> sector_procedencia=0 66 <conf:(0.9)> lift:(1.03) lev:(
195. genero=0 sector_procedencia=0 73 ==> trabaja=0 66 <conf:(0.9)> lift:(0.99) lev:(-0) [0] conv:(0.81)
196. genero=0 tipo_colegio=2 73 ==> sector_procedencia=0 66 <conf:(0.9)> lift:(1.03) lev:(0.02) [1] conv:(1.0
197. genero=0 sector_procedencia=0 73 ==> tipo_colegio=2 66 <conf:(0.9)> lift:(0.99) lev:(-0) [0] conv:(0.81)
198. provincia_nacimiento=11 trabaja=0 CALCULO_DIFERENCIAL=1 52 ==> sector_procedencia=0 47 <conf:(0.9)> lift
199. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 52 ==> sector
200. provincia_nacimiento=11 trabaja=0 CALCULO_DIFERENCIAL=1 52 ==> tipo_colegio=2 47 <conf:(0.9)> lift:(0.99
  
```

Figura 256. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza mayor a 98% y lift mayor a 1 respectivamente, se obtiene un total de 16 reglas (Ver Figura 3).

REGLAS	CONF	LIFT	
1. provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 49 ==> CALCULO_DIFERENCIAL=1 49	conf:(1)>	lift:(1.31)	lev:(0.12)
2. provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 ==> CALCULO_DIFERENCIAL=1 47	conf:(1)>	lift:(1.31)	lev:(0.11)
3. sector_procedencia=0 provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 44 ==> CALCULO_DIFERENCIAL=1 44	conf:(1)>	lift:(1.31)	lev:(0.1)
4. provincia_nacimiento=11 trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 43 ==> CALCULO_DIFERENCIAL=1 43	conf:(1)>	lift:(1.31)	lev:(0.1)
5. sector_procedencia=0 provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 42 ==> CALCULO_DIFERENCIAL=1 42	conf:(1)>	lift:(1.31)	lev:(0.1)
6. provincia_nacimiento=11 tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 41 ==> CALCULO_DIFERENCIAL=1 41	conf:(1)>	lift:(1.31)	lev:(0.1)
7. provincia_nacimiento=11 trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 41 ==> CALCULO_DIFERENCIAL=1 41	conf:(1)>	lift:(1.31)	lev:(0.1)
48. provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 49 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47	conf:(0.96)>	lift:(1.4)	lev:(0.13)
61. sector_procedencia=0 provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 44 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 41	conf:(0.95)>	lift:(1.4)	lev:(0.12)
69. provincia_nacimiento=11 trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 43 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 41	conf:(0.95)>	lift:(1.4)	lev:(0.12)
71. PROGRAMACION_ORIENTADA_A_OBJETOS=1 62 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 59	conf:(0.95)>	lift:(1.39)	lev:(0.16)
75. sector_procedencia=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 55 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 52	conf:(0.95)>	lift:(1.38)	lev:(0.14)
76. trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 55 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 52	conf:(0.95)>	lift:(1.38)	lev:(0.14)
79. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 54 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 51	conf:(0.94)>	lift:(1.38)	lev:(0.14)
83. genero=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 50 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47	conf:(0.94)>	lift:(1.38)	lev:(0.13)
84. trabaja=0 tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 50 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47	conf:(0.94)>	lift:(1.38)	lev:(0.13)

Figura 257. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:



- 1) Estudiante oriundo de la Provincia de Loja y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 2) Estudiante oriundo de la Provincia de Loja y que apruebe las materias de Programación Orientada a Objetos y Teoría de la Distribución y Probabilidad, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 3) Estudiante perteneciente al sector Urbano, oriundo de la Provincia de Loja y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 4) Si un estudiante oriundo de la Provincia de Loja, no ejerce actividad laboral y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 5) Estudiante perteneciente al sector Urbano, oriundo de la Provincia de Loja y que apruebe las materias de Programación Orientada a Objetos y Teoría de la Distribución y Probabilidad, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 6) Estudiante oriundo de la Provincia de Loja, que haya culminado sus estudios secundarios en una Institución Fiscomisional y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 7) Estudiante oriundo de la provincia de Loja, no ejerce actividad laboral y que apruebe las materias de Programación Orientada a Objetos y Teoría de la Distribución y Probabilidad, es muy probable que también apruebe a la asignatura de Cálculo Diferencial.
- 8) Estudiante oriundo de la Provincia de Loja y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 9) Estudiante perteneciente al sector Urbano, oriundo de la Provincia de Loja y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

- 10) Si un estudiante oriundo de la Provincia de Loja, no ejerce actividad laboral y aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 11) Estudiante aprueba la materia de Programación Orientada a Objetos es muy probable que también apruebe las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 12) Estudiante perteneciente al sector Urbano y que apruebe la materia de Programación Orientada a Objetos, es muy probable que también apruebe a la asignatura de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 13) Si un estudiante no ejerce actividad laboral y aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe a las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 14) Si un estudiante culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe a las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 15) Si un estudiante no ejerce actividad laboral y que aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe a las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 16) Si un estudiante no ejerce actividad laboral, culmina sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe a las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.

Para este ciclo, la correlación entre atributos es muy elevada, tanto en la confianza como en el lift. Se muestran casos en los que el lift obtiene valores elevados (Ej. Regla #48 con lift de 1.4), pero no una confianza al 100%, pero esto no indica que el nivel de confiabilidad sea malo, ya que en reglas como la mencionada anteriormente se tiene un 96% de fiabilidad, es decir, valores más



reales y predictivos. Por otra parte, se han determinado Reglas definiendo lift como predeterminado, obteniendo valores mayores a 1.40 y una confiabilidad de las premisas del 68% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS		CONF	LIFT	LEV
1. provincia_nacimiento=11 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 53 ==> PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.89)	lift:(1.44)>	lev:(0.14)	
2. PROGRAMACION_ORIENTADA_A_OBJETOS=1 62 ==> provincia_nacimiento=11 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 con	:(0.76)	lift:(1.44)>	lev:(0.14)	
3. tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 54 ==> trabaja=0 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 con	:(0.87)	lift:(1.44)>	lev:(0.14)	
4. trabaja=0 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 61 ==> tipo_colegio=2 PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.77)	lift:(1.44)>	lev:(0.14)	
5. trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 55 ==> tipo_colegio=2 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 con	:(0.85)	lift:(1.44)>	lev:(0.14)	
6. tipo_colegio=2 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 60 ==> trabaja=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.78)	lift:(1.44)>	lev:(0.14)	
7. provincia_nacimiento=11 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 55 ==> PROGRAMACION_ORIENTADA_A_OBJETOS=1 CALCULO_DIFERENCIAL=1 47 con	:(0.85)	lift:(1.41)>	lev:(0.14)	
8. PROGRAMACION_ORIENTADA_A_OBJETOS=1 CALCULO_DIFERENCIAL=1 61 ==> provincia_nacimiento=11 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 con	:(0.77)	lift:(1.41)>	lev:(0.14)	
9. sector_procedencia=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 55 ==> tipo_colegio=2 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 60 ==> sector_procedencia=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.84)	lift:(1.41)>	lev:(0.13)	
10. tipo_colegio=2 CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 60 ==> sector_procedencia=0 PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.77)	lift:(1.41)>	lev:(0.13)	
11. CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 69 ==> provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 47 con	:(0.68)	lift:(1.4)>	lev:(0.13)	
12. provincia_nacimiento=11 PROGRAMACION_ORIENTADA_A_OBJETOS=1 49 ==> CALCULO_DIFERENCIAL=1 TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD=1 47 con	:(0.96)	lift:(1.4)>	lev:(0.13)	

Figura 258. Reglas obtenidas en Weka con Lift mayor a 1

- 13) Estudiante oriundo de la Provincia de Loja y que aprueba las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad, es muy probable que también la materia de Programación Orientada a Objetos.
- 14) Estudiante que apruebe la asignatura de Programación Orientada a Objetos es muy probable que sea del sector Urbano y que también apruebe las materias de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 15) Si un estudiante ha culminado sus estudios secundarios en una Institución Fiscomisional y aprueba la materia de Programación Orientada a Objetos, es muy probable que no cuente con ninguna actividad laboral y que también apruebe las materias de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 16) Si un estudiante no ejerce ninguna actividad laboral y aprueba las materias de Cálculo Diferencial y Teoría de la Distribución y Probabilidad, es muy probable que haya culminado sus estudios secundarios en una Institución Fiscomisional y que también apruebe la materia de Programación Orientada a Objetos.
- 17) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Programación Orientada a Objetos, es muy probable que sus estudios secundarios los haya culminado en una Institución Fiscomisional y que apruebe las materias de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.



- 18) Estudiante con estudios secundarios culminados en una Institución Fiscomisional y aprueba la materia de Cálculo Diferencial y Teoría de la Distribución y Probabilidad.
- 19) Estudiante oriundo de la provincia de Loja y que aprueba la materia de Teoría de la Distribución y Probabilidad, es muy probable que apruebe también las materias de Programación Orientada a Objetos y Cálculo Diferencial.
- 20) Estudiante que pruebe la asignatura de Programación Orientada a Objetos y Cálculo Diferencial, es muy probable que pertenezca a la Provincia de Loja y que apruebe la materia de Teoría de la Distribución y Probabilidad.
- 21) Estudiante procedente del sector Urbano y aprueba la materia de Programación Orientada a Objetos, es muy probable que sus estudios secundarios los haya culminado en una Institución Fiscomisional y apruebe las materias de Cálculo diferencial y Teoría de la Distribución y Probabilidad.
- 22) Si un estudiante culmina sus estudios secundarios en una Institución Fiscomisional y aprueba las materias de Cálculo diferencial y Teoría de la Distribución y Probabilidad, es muy probable que pertenezca al sector Urbano y aprueba la materia de Programación Orientada a Objetos.
- 23) Si un estudiante aprueba las materias de Cálculo diferencial y Teoría de la Distribución y Probabilidad, es muy probable que sea oriundo de la provincia de Loja y que también apruebe la materia de Programación Orientada a Objetos.
- 24) Estudiante oriundo de la provincia de Loja y aprueba la materia de Programación Orientada a Objetos, es muy probable que apruebe las materias de Cálculo diferencial y Teoría de la Distribución y Probabilidad.

De las premisas obtenidas, considerando el lift como métrica principal, se identifica a reglas con valor máximo de 1.44 y la mínima con 1.4, sin embargo, estas reglas con alto lift, muestran una baja confiabilidad en las mismas. Por ejemplo, las reglas #4 y 6 tienen un lift alto pero la confianza es del 77 y 78% respectivamente, es decir, si bien es cierto tendrán un alto valor predictivo y aunque su confianza no es la mejor, en estos casos es beneficioso que las





premisas tengan una buena confiabilidad. Por otra parte, la regla #1 menciona que si un estudiante es oriundo de la Provincia de Loja y aprueba las asignaturas de Cálculo Diferencial y Teoría de la Distribución y Probabilidad, es muy probable que también apruebe la materia de Programación Orientada a Objetos, a esta regla se suma la confiabilidad del 89%, la cual ya podría ser considerada como una base para la toma de decisiones. De la misma manera, se identifica a la regla # 12 con mayor confiabilidad con un valor del 96% y un lift de 1.4, la cual describe: si estudiante oriundo de la provincia de Loja, aprueba la materia de Programación Orientada a Objetos, es muy probable que también apruebe las materias de Cálculo diferencial y Teoría de la Distribución y Probabilidad.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, es decir, guardan información relacionada entre sí y con las materias relacionadas con las matemáticas, ante ello se puede identificar que las características de no trabajar, de ser nativo de la provincia de Loja, de culminar sus estudios secundarios en un colegio Fiscomisional y de formar parte del sector Urbano están muy presentes dentro de las premisas obtenidas. De la misma manera, la relación de las premisas con el género femenino en particular es bajo, principalmente en estos ciclos no se identifica la presencia de las mismas.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en Weka. La relación entre las asignaturas se sigue manteniendo, se muestra la presencia de estudiante provenientes de la Provincia de Loja, además, en este caso se muestra la característica tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 32 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

### AssociationRules

```
Association Rules
[COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, provincia_nacimiento, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [edad, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[tipo_colegio, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, tipo_colegio, COMUNICACION PROFESIONAL] --> [INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[COMUNICACION PROFESIONAL] --> [edad, provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[edad, COMUNICACION PROFESIONAL] --> [provincia_nacimiento, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
[provincia_nacimiento, COMUNICACION PROFESIONAL] --> [edad, tipo_colegio, INTRODUCCION A LAS CIENCIAS DE LA COMPUTACION] (confidence: 0.803)
```

Figura 259. Reglas de asociación en RapidMiner

### Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se logra identificar que, la relación entre las asignaturas con las variables demográficas o sociales es un poco escasa, ya que el nivel de confiabilidad en las mismas no alcanza el 70 %, mientras tanto que, la relación únicamente en variables sociales y demográficas es muy alta, estas premisas tienen una confianza de hasta el 90%, con una predicción futura del 1.23. Las características que se identifican con más periodicidad dentro de las premisas son aquellos estudiantes de género masculino, provenientes del sector Urbano, de la provincia de Loja y de una Institución Privada, son solteros y no cuentan con ninguna actividad laboral, estas son las características más relevantes que se destacan dentro de las premisas.



```
Apriori_Final.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se guardaron todos los cambios
+ Código + Texto
RAM Disco Editando
RelationRecord(items=frozenset({'TDP_A', 'CD_A'}), support=0.6831683168316832, ordered_statistics=[OrderedStatistic(items_base=frozenset({'CD_A'}), items_add=fr
RelationRecord(items=frozenset({'Urbano', 'CD_A'}), support=0.6831683168316832, ordered_statistics=[OrderedStatistic(items_base=frozenset({'CD_A'}), items_add=
RelationRecord(items=frozenset({'Masculino', 'Urbano'}), support=0.7227272727272727, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Masculino'}), :
RelationRecord(items=frozenset({'NTrabaja', 'privado'}), support=0.8613861386138614, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'}), it
RelationRecord(items=frozenset({'TDP_A', 'Urbano'}), support=0.6534653465346535, ordered_statistics=[OrderedStatistic(items_base=frozenset({'TDP_A'}), items_ad
RelationRecord(items=frozenset({'privado', 'Urbano'}), support=0.8118811881188119, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Urbano'}), items
RelationRecord(items=frozenset({'11', 'Masculino', 'Soltero'}), support=0.6138613861386139, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero
RelationRecord(items=frozenset({'11', 'NTrabaja', 'privado'}), support=0.6435643564356436, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja
RelationRecord(items=frozenset({'11', 'privado', 'Urbano'}), support=0.6138613861386139, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Urbano'}),
RelationRecord(items=frozenset({'NTrabaja', 'TDP_A', 'CD_A'}), support=0.6039603960396039, ordered_statistics=[OrderedStatistic(items_base=frozenset({'CD_A'}),
RelationRecord(items=frozenset({'NTrabaja', 'Urbano', 'CD_A'}), support=0.6039603960396039, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Urbano'
RelationRecord(items=frozenset({'NTrabaja', 'privado', 'CD_A'}), support=0.6237623762376238, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTraba
```

Figura 260. Reglas de asociación obtenidas por Python

## Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en Weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

### Clustered Instances

0	16 ( 16%)
1	40 ( 40%)
2	26 ( 26%)
3	10 ( 10%)
4	9 ( 9%)

Figura 261. Clústers con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 1, y 2. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

Final cluster centroids:

Attribute	Full Data (101.0)	Cluster#				
		0 (16.0)	1 (40.0)	2 (26.0)	3 (10.0)	4 (9.0)
edad	20	21	20	21	21	22
genero	0	0	0	0	1	0
sector_procedencia	0	1	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	0	0	0	0	0
tipo_colegio	2	2	2	2	2	2
PROGRAMACION_ORIENTADA_A_OBJETOS	1	1	1	0	1	1
CALCULO_DIFERENCIAL	1	1	1	0	1	1
TEORIA_DE_LA_DISTRIBUCION_Y_PROBABILIDAD	1	1	1	0	1	1

Figura 262. Clústers obtenidos por Weka.

**Clúster\_0 (16%):** Estudiantes que tengan una edad de 21 años, pertenezcan al género masculino, sean del sector Rural y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, estas son: Programación Orientada a Objetos, Cálculo Diferencial y Teoría de la Distribución y Probabilidad.

**Clúster\_1 (40%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios



secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, estas son: Programación Orientada a Objetos, Cálculo Diferencial y Teoría de la Distribución y Probabilidad.

**Clúster\_2 (26%):** Estudiantes que tengan una edad de 21 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, reprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, estas son: Programación Orientada a Objetos, Cálculo Diferencial y Teoría de la Distribución y Probabilidad.

**Clúster\_3 (10%):** Estudiantes que tengan una edad de 21 años, pertenezcan al género femenino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, estas son: Programación Orientada a Objetos, Cálculo Diferencial y Teoría de la Distribución y Probabilidad.


**Clúster\_4 (9%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al segundo ciclo, estas son: Programación Orientada a Objetos, Cálculo Diferencial y Teoría de la Distribución y Probabilidad.

### **Resultados obtenidos por Algoritmo K - means – RapidMiner**

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústers.

Nominal values

×



Index	Nominal value	Absolute count	Fraction
1	cluster_0	58	0.574
2	cluster_1	27	0.267
3	cluster_2	7	0.069
4	cluster_4	5	0.050
5	cluster_3	4	0.040

Figura 263. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 101 instancias están clasificadas o agrupadas a su clúster más cercano.

Open in Turbo Prep Auto Model Filter (101 / 101 examples): all

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	PROGRAMA...	CALCULO DL...	TEORIA DE L...	tipo
1	1	cluster_2	20	0	0	7	0	0	1	1	0.16
2	2	cluster_4	23	0	0	11	0	1	1	1	0.16
3	3	cluster_1	34	1	0	7	0	0	0	0	0.16
4	4	cluster_0	21	0	0	11	0	1	1	1	0.16
5	5	cluster_0	20	0	0	11	0	0	0	0	0.16
6	6	cluster_3	20	0	0	19	0	1	1	1	0.16
7	7	cluster_0	19	0	0	11	0	0	1	1	0.16
8	8	cluster_4	23	0	0	11	0	1	1	1	0.16
9	9	cluster_0	21	0	0	10	0	0	1	1	0.16
10	10	cluster_0	20	0	0	11	0	0	0	1	0.16
11	11	cluster_4	27	0	0	11	1	1	1	1	3.54
12	12	cluster_0	20	0	0	11	0	0	1	1	0.16
13	13	cluster_0	20	0	0	11	0	0	1	1	0.16

Figura 264. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 0 con más instancias agrupadas.



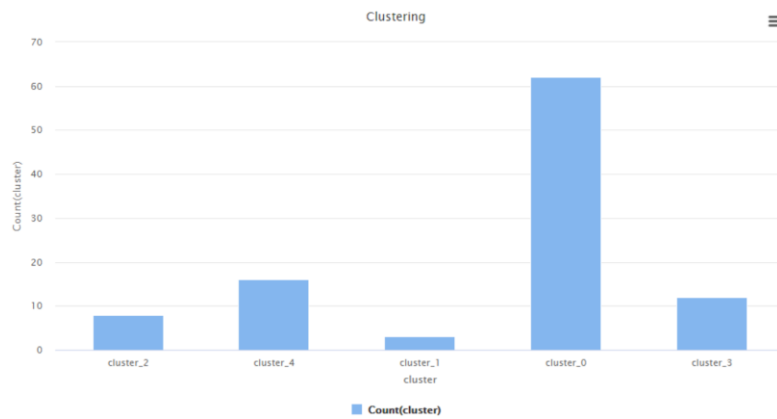


Figura 265. Representación gráfica de los clústeres obtenidos

Por otra parte, haciendo uso de una de las alternativas de RapidMiner como es Auto Model, donde como su nombre lo indica, realiza un modelo automático únicamente cargando el conjunto de datos y seleccionando que tipo de algoritmo que se desea ejecutar, con ello se logró obtener las siguientes agrupaciones, cada una de ellas con sus respectivas características y valores correspondientes.

#### k-Means - Centroid Table

Cluster	CALCULO DI...	edad	genero	PROGRAMA...	provincia_n...	sector_proc...	TEORIA DE L...	tipo_colegio	trabaja
Cluster 0	0.154	21.577	0.115	-0.000	12.231	0.154	0.115	2	0.038
Cluster 1	1.000	20.786	1	0.786	10.429	0.143	0.857	2	-0
Cluster 2	1	25.286	-0	0.857	11	0.286	1	1	0.429
Cluster 3	0.500	34.000	1	0.500	9	0	0.500	2	0.500
Cluster 4	0.981	21.577	0.000	0.846	11.827	0.077	0.962	2.038	0.077

Figura 266. Clústeres obtenidos por Auto Model dentro de Rapidminer

**Clúster 0:** Estudiante masculino procedente de la provincia de Los Ríos, que reside en el sector Urbano, no ejerce ninguna actividad laboral y culmina sus estudios secundarios en una Institución Fiscomisional, tiene una probabilidad mínima de aprobar a las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Programación Orientada a Objetos y Teoría de la Distribución y la Probabilidad.

**Clúster 1:** Estudiante de género femenino procedente de la provincia de Loja, que reside en el sector Urbano, no ejerce ninguna actividad laboral y culmina sus estudios secundarios en una Institución Fiscomisional, aprobará la materia de Cálculo Diferencial y también posee una probabilidad alta de aprobar a las



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

asignaturas de Programación Orientada a Objetos y Teoría de la Distribución y la Probabilidad.

**Clúster 2:** Estudiante masculino procedente de la provincia de Loja, que reside en el sector Urbano, cuenta con una actividad laboral y culmina sus estudios secundarios en una Institución Fiscal, aprobará las materias de Cálculo Diferencial y Teoría de la Distribución y la Probabilidad y con una probabilidad de aprobar a la asignatura de Programación Orientada a Objetos.

**Clúster 3:** Estudiante de género femenino procedente de la provincia del Guayas, que reside en el sector Urbano, no ejerce ninguna actividad laboral y culmina sus estudios secundarios en una Institución Fiscomisional, tiene una probabilidad baja de aprobar a las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Programación Orientada a Objetos y Teoría de la Distribución y la Probabilidad.

**Clúster 4:** Estudiante masculino procedente de la provincia de Loja, que reside en el sector Urbano, no ejerce ninguna actividad laboral y culmina sus estudios secundarios en una Institución Fiscomisional, tiene una probabilidad muy alta de aprobar a las asignaturas relacionadas a las matemáticas como Cálculo Diferencial, Programación Orientada a Objetos y Teoría de la Distribución y la Probabilidad.

### **Resultados obtenidos por Algoritmo K - means – Python**

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide rojo y verde, se observa una relación entre los estudiantes de género masculino y femenino ambos con

una Institución Fiscomisional, mientras que en el centroide de color azul hace referencia a estudiantes de género masculino con un colegio Fiscal.

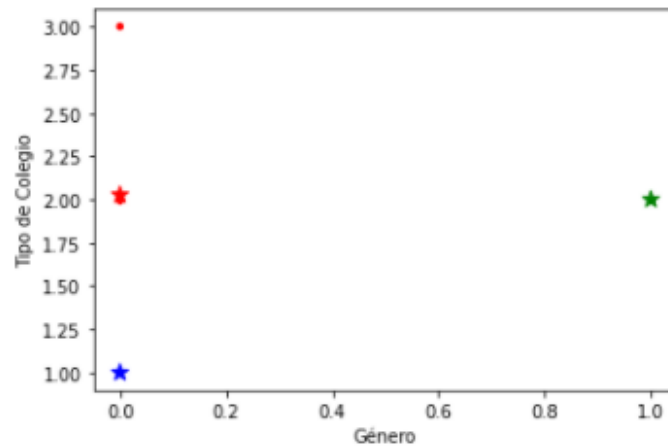


Figura 267. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Programación Orientada a Objetos** en donde los estudiantes con edades aproximadas a 21, 24 y 33 años aprobarán las materias de manera exitosa.

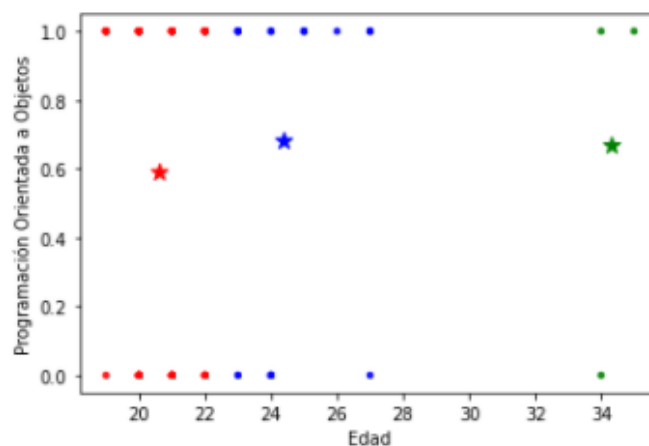


Figura 268. Relación Edad – Ingeniería de Software II

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género y Programación Orientada a Objetos** donde, se observa que un grupo de estudiantes pertenecientes al género femenino tienen una probabilidad aproximada del 60% de aprobar la materia, por otro lado, el género masculino si aprobará con éxito a la misma, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprueban la asignatura de Programación Orientada a Objetos.



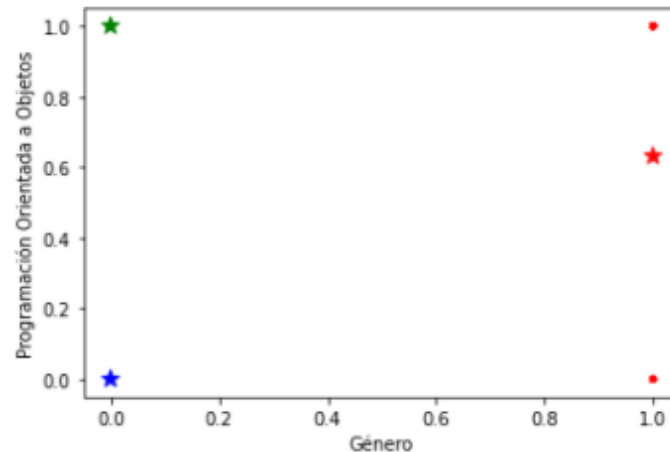


Figura 269. Relación Género – Ingeniería de Software II

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Cálculo Diferencial** y **Programación Orientada a Objetos**, en el clúster verde se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde apruebe Cálculo Diferencial pero reprueba la materia de Programación Orientada a Objetos (clúster rojo), y por último dentro del clúster rojo se muestra una baja probabilidad que el estudiante únicamente apruebe cualquiera de las dos asignaturas.

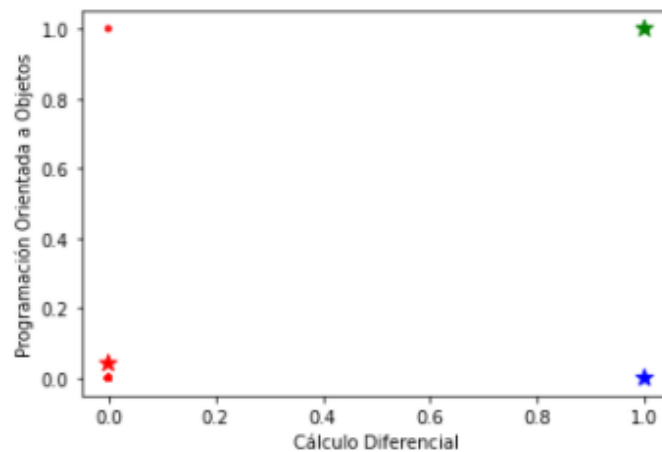


Figura 270. Relación Modelamiento Matemático – Ingeniería de Software II

## Anexo 16. Ingeniería en Computación – Tercer Ciclos

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **tercer ciclo de la carrera de Ingeniería en Computación** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas Weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

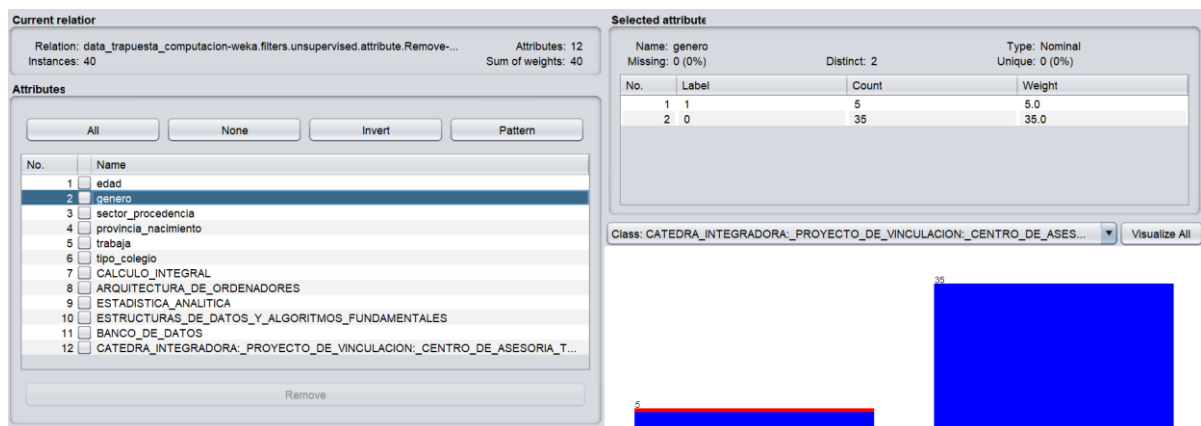


Figura 271. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.

```
Associator output
178. trabaja=0 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05)
179. trabaja=0 34 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.08)
180. ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 34 ==> CALCULO_INTEGRAL=1 33 <conf:(0.97)> lift:(1) lev:(-0) [0] conv:(0.42)
181. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
182. ESTADISTICA_ANALITICA=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> CALCULO_INTEGRAL=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
183. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 34 ==> CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
184. ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 34 ==> ARQUITECTURA_DE_ORDENADORES=1 33 <conf:(0.97)> lift:(1.11) lev:(0.08) [3] conv:(2.13)
185. ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 34 ==> ESTADISTICA_ANALITICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
186. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
187. ESTADISTICA_ANALITICA=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> ARQUITECTURA_DE_ORDENADORES=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
188. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 34 ==> CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
189. ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 34 ==> ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
190. ARQUITECTURA_DE_ORDENADORES=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
191. ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 34 ==> CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
192. ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 34 ==> ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
193. ESTADISTICA_ANALITICA=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
194. ESTADISTICA_ANALITICA=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
195. ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 34 ==> CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
196. genero=0 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
197. genero=0 CALCULO_INTEGRAL=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
198. genero=0 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 34 ==> CALCULO_INTEGRAL=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
199. genero=0 CALCULO_INTEGRAL=1 34 ==> ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
200. genero=0 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 34 ==> BANCO_DE_DATOS=1 33 <conf:(0.97)> lift:(1.05) lev:(0.04) [1] conv:(1.27)
```

Figura 272. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza del 100% y lift mayor a 1 respectivamente, se obtiene un total de 17 reglas (Ver Figura 3).

REGLAS	CONF	LIFT
41. trabaja=0 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33	conf(1)	lift(1.11)
53. ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33	conf(1)	lift(1.11)
57. ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 33 ==> CALCULO_INTEGRAL=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 33	conf(1)	lift(1.11)
87. trabaja=0 CALCULO_INTEGRAL=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
91. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> CALCULO_INTEGRAL=1 BANCO_DE_DATOS=1 32	conf(1)	lift(1.11)
92. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 32	conf(1)	lift(1.18)
97. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> ARQUITECTURA_DE_ORDENADORES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
103. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
107. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
113. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> CALCULO_INTEGRAL=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32	conf(1)	lift(1.11)
114. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> CALCULO_INTEGRAL=1 BANCO_DE_DATOS=1 32	conf(1)	lift(1.11)
116. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> ARQUITECTURA_DE_ORDENADORES=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32	conf(1)	lift(1.18)
117. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 32	conf(1)	lift(1.18)
118. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 BANCO_DE_DATOS=1 ==> ARQUITECTURA_DE_ORDENADORES=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
119. CALCULO_INTEGRAL=1 ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.11)
120. ARQUITECTURA_DE_ORDENADORES=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> CALCULO_INTEGRAL=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.14)
121. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA=1 32	conf(1)	lift(1.21)

Figura 273. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia Estructura de Datos y Algoritmos Fundamentales, es muy probable que también apruebe las materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 2) Si un estudiante aprueba las materias de Estadística Analítica y Estructura de Datos y algoritmos Fundamentales, es muy probable que apruebe las



- materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 3) Si un Estudiante aprueba la materia de Arquitectura de Computadores, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que también apruebe las materias de Cálculo Integral y Estructura de Datos y Algoritmos Funcionales.
  - 4) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Cálculo Integral y Estructura de Datos y Algoritmos Funcionales, es muy probable que también apruebe las materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
  - 5) Si un estudiante aprueba las materias de Arquitectura de Computadores, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que apruebe las materias de Cálculo Integral y Banco de Datos.
  - 6) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que apruebe las materias de Arquitectura de Computadores y Banco de Datos.
  - 7) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que apruebe las materias de Arquitectura de Computadores y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
  - 8) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que apruebe las materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
  - 9) Si un estudiante aprueba las materias de Arquitectura de Computadores, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que apruebe las materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
  - 10) Si un estudiante aprueba las materias de Arquitectura de Computadores, Estadística Analítica, Banco de Datos y Catedra Integradora: Proyecto de



Vinculación: "Centro de Asesoría Tecnológica", es muy probable que apruebe las materias de Cálculo Integral y Estructura de Datos y Algoritmos Fundamentales.

- 11) Si un estudiante aprueba las materias de Arquitectura de Computadores, Estadística Analítica, Estructura de Datos y Algoritmos Funcionales y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que apruebe las materias de Cálculo Integral y Banco de Datos.
- 12) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que apruebe las materias de Arquitectura de Computadores y Estructura de Datos y Algoritmos Funcionales.
- 13) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica, Estructura de Datos y Algoritmos Funcionales y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que apruebe las materias de Arquitectura de Computadores y Banco de Datos.
- 14) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica, Estructura de Datos y Algoritmos Funcionales y Banco de datos, es muy probable que también apruebe las materias de Arquitectura de Ordenadores y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 15) Si un estudiante aprueba las materias de Cálculo Integral, Arquitectura de Computadores, Estadística Analítica, Estructura de Datos y Algoritmos Funcionales y Banco de datos, es muy probable que también apruebe las materias de Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 16) Si un estudiante aprueba las materias de Arquitectura de Computadores, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que también apruebe las materias de Cálculo Integral, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".

17) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica, y Estructura de Datos y Algoritmos Funcionales, es muy probable que también apruebe las materias de Arquitectura de Computadores, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".

Dentro de las reglas obtenidas se identifica la fuerte correlación que existen entre la confianza y el lift, todas las premisas cuentan con un valor de confiabilidad del 100% y de la misma manera todas ellas cuentan con un lift, donde pueden ser consideradas para valores predictivos, una de las reglas con valores altos es la # 92, misma que refleja a la relación entre las materias impartidas durante el periodo académico, por ejemplo, si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Funcionales, es muy probable que el alumno también apruebe las materias de Arquitectura de Computadores y Banco de Datos. Por otro lado, la regla # 121, indica a la aprobación de las materias Arquitectura de Computadores, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", si el estudiante aprueba las asignaturas de Cálculo Integral, Estadística Analítica, y Estructura de Datos y Algoritmos Funcionales. En este ciclo particularmente, únicamente se muestra la característica de trabajo y más relación entre asignaturas pertenecientes al mismo.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado, obteniendo valores mayores a 1.25 y una confiabilidad de las premisas del 84% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONF	LIFT
1. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 30 com	{0.94}	lift:(1.25)>
2. trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 30 ==> CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 30 com	{1}	lift:(1.25)>
3. trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 30 ==> CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 :{1}	lift:(1.25)>	
4. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 32 ==> trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 :{0.94}	lift:(1.25)>	
5. trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 30 ==> CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 :{1}	lift:(1.25)>	
6. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 :{0.94}	lift:(1.25)>	
7. trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 30 ==> CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 :{1}	lift:(1.25)>	
8. CALCULO_INTEGRAL=1 ESTADISTICA_ANALITICA=1 BANCO_DE_DATOS=1 CATEDRA_INTEGRADORA: PROYECTO_DE_VINCULACION: CENTRO_DE_ASESORIA_TECNOLOGICA=1 32 ==> trabaja=0 ARQUITECTURA_DE_ORDENADORES=1 ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES=1 :{0.94}	lift:(1.25)>	

Figura 274. Reglas obtenidas en Weka con Lift mayor a 1

1) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Fundamentales, es muy probable que no ejerza ninguna actividad laboral y apruebe las asignaturas de Arquitectura de Ordenadores y Banco de Datos.





UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

- 2) Si un estudiante no cuenta con ninguna actividad laboral y aprueba las materias de Arquitectura de Ordenadores, Banco de Datos, es muy probable que también apruebe las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Fundamentales.
- 3) Estudiante que no ejerce ninguna actividad laboral y aprueba las materias de Arquitectura de Ordenadores, Estadística Analítica y Algoritmos Fundamentales, es muy probable que también apruebe las materias de Cálculo Integral, Estadística Analítica, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 4) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Fundamentales, es muy probable que no ejerza ninguna actividad laboral y que también apruebe a las asignaturas de Arquitectura de Ordenadores, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 5) Si un estudiante no ejerce ninguna actividad laboral y aprueba las materias de Arquitectura de Ordenadores y Banco de Datos, es muy probable que apruebe también las asignaturas de Cálculo Integral, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales y Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica".
- 6) Si un estudiante aprueba las materias de Cálculo Integral, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales y Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que no trabaje y que apruebe las asignaturas de Arquitectura de Ordenadores y Banco de Datos.
- 7) Si un estudiante no está laborando y aprueba las materias de Arquitectura de Ordenadores, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que apruebe las asignaturas de Cálculo Integral, Estadística Analítica y Estructura de Datos y Algoritmos Fundamentales.
- 8) Cálculo Integral, Estadística Analítica, Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica", es muy probable que no trabaje y que también apruebe las materias de



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

## Arquitectura de Ordenadores y Estructura de Datos y Algoritmos Fundamentales.

Cada una de las reglas obtenidas muestran la relación entre las variables demográficas, sociales y académicas, así como también, las materias relacionadas con las matemáticas. En virtud de ello, se puede identificar que la característica de no trabajar predomina en todas las premisas obtenidas, considerando que, si bien es cierto dentro de la carrera en estudio los estudiantes no cuentan con ninguna actividad laboral, esto en relación a pertenecer a un colegio Fiscomisional. En el mismo contexto, tanto la confiabilidad como la predicción en las premisas obtenidas son muy buenas, ya que se cuenta con una confianza mínima del 94% y un lift de 1.25, esto quiere decir que si un estudiante no ejerce ninguna actividad laboral y aprueba las materias de Arquitectura de Ordenadores y Banco de Datos, la probabilidad de que apruebe también las asignaturas de Cálculo Integral, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales y Banco de Datos y Catedra Integradora: Proyecto de Vinculación: "Centro de Asesoría Tecnológica" es muy alta (Ver regla # 5). De la misma manera, la relación de las premisas con el género femenino en particular mínima, esto se debe a que como es de conocimiento general el número de mujeres presentes en este tipo de Ingenierías es mínimo, además de que, como se refleja en las premisas se relaciona más con variables demográficas y sociales en lugar de las académicas.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en Weka. La relación entre las asignaturas se sigue manteniendo, se muestra la presencia de estudiante provenientes de la Provincia de Loja, además, en este caso se muestra la característica tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 32 años, la mayoría de edad se debe



a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

## AssociationRules

Association Rules

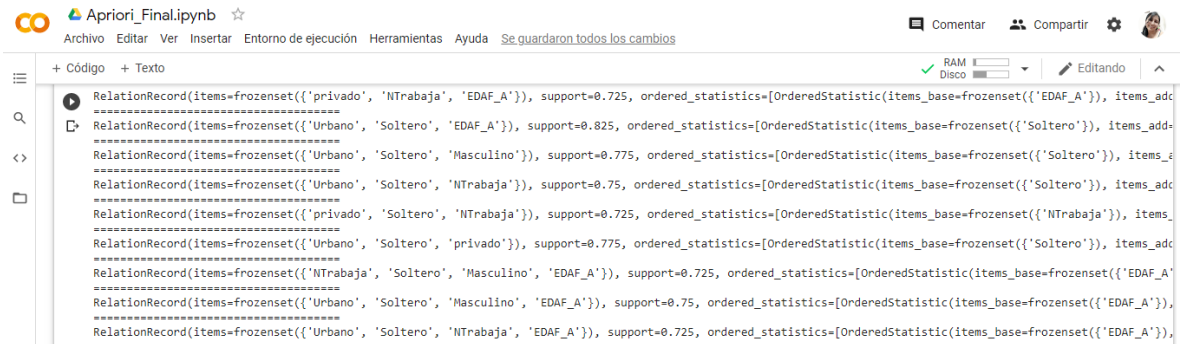
```
[provincia_nacimiento] --> [tipo_colegio, CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", BAN
[tipo_colegio] --> [provincia_nacimiento, CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", BAN
[provincia_nacimiento, tipo_colegio] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", BAN
[provincia_nacimiento] --> [tipo_colegio, CALCULO INTEGRAL, estado_civil, BANCOS DE DATOS] (confidence: 0.875)
[tipo_colegio] --> [provincia_nacimiento, CALCULO INTEGRAL, estado_civil, BANCOS DE DATOS] (confidence: 0.875)
[provincia_nacimiento, tipo_colegio] --> [CALCULO INTEGRAL, estado_civil, BANCOS DE DATOS] (confidence: 0.875)
[provincia_nacimiento] --> [tipo_colegio, CALCULO INTEGRAL, estado_civil, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[tipo_colegio] --> [provincia_nacimiento, CALCULO INTEGRAL, estado_civil, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento, tipo_colegio] --> [CALCULO INTEGRAL, estado_civil, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento] --> [tipo_colegio, CALCULO INTEGRAL, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[tipo_colegio] --> [provincia_nacimiento, CALCULO INTEGRAL, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento, tipo_colegio] --> [CALCULO INTEGRAL, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento] --> [tipo_colegio, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ARQUITE
[tipo_colegio] --> [provincia_nacimiento, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ARQUITE
[provincia_nacimiento, tipo_colegio] --> [CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ARQUITE
[provincia_nacimiento] --> [tipo_colegio, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, BANCOS
[tipo_colegio] --> [provincia_nacimiento, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, BANCOS
[provincia_nacimiento, tipo_colegio] --> [CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, BANCOS
[provincia_nacimiento] --> [tipo_colegio, estado_civil, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[tipo_colegio] --> [provincia_nacimiento, estado_civil, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento, tipo_colegio] --> [estado_civil, BANCOS DE DATOS, ESTRUCTURAS DE DATOS Y ALGORITMOS FUNDAMENTALES] (confidence: 0.875)
[provincia_nacimiento] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ARC
[provincia_nacimiento] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, EST
[provincia_nacimiento] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, EST
[provincia_nacimiento] --> [CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, BANCOS DE DATOS, ESTE
[tipo_colegio] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ARQUITECTUR
[tipo_colegio] --> [CALCULO INTEGRAL, CATEDRA INTEGRADORA: PROYECTO DE VINCULACION: "CENTRO DE ASESORIA TECNOLÓGICA", estado_civil, ESTRUCTURAS
```

Figura 275. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se logra identificar que, la relación entre las diferentes asignaturas relacionadas con las matemáticas con las variables sociales o demográficas ya es visible, para estas premisas la confiabilidad supera al 70%, con una predicción a futuro de hasta 1.23. La asignatura que sobresale es la de Estructura de Datos y Algoritmos Funcionales, esto en relación con estudiantes procedentes del sector Urbano, que no ejerzan actividad laboral, sean solteros y

provenientes de una Institución Fiscomisional, estas son las características más relevantes que se destacan dentro de las premisas.



```

+ Código + Texto
RAM
Disco
Editando
RelationRecord(items=frozenset({'privado', 'NTrabaja', 'EDAF_A'}), support=0.725, ordered_statistics=[OrderedStatistic(items_base=frozenset({'EDAF_A'}), items_add
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'EDAF_A'}), support=0.825, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'}), items_add
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'Masculino'}), support=0.775, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'}), items_e
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'NTrabaja'}), support=0.75, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'}), items_adc
RelationRecord(items=frozenset({'privado', 'Soltero', 'NTrabaja'}), support=0.725, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'}), items_
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'privado'}), support=0.775, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'}), items_adc
RelationRecord(items=frozenset({'NTrabaja', 'Soltero', 'Masculino', 'EDAF_A'}), support=0.725, ordered_statistics=[OrderedStatistic(items_base=frozenset({'EDAF_A'
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'Masculino', 'EDAF_A'}), support=0.75, ordered_statistics=[OrderedStatistic(items_base=frozenset({'EDAF_A'}),
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'NTrabaja', 'EDAF_A'}), support=0.725, ordered_statistics=[OrderedStatistic(items_base=frozenset({'EDAF_A'})

```

Figura 276. Reglas de asociación obtenidas por Python

### Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

#### Clustered Instances

0	6 ( 15%)
1	19 ( 48%)
2	8 ( 20%)
3	3 ( 8%)
4	4 ( 10%)

Figura 277. Clústers con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 0, 1, y 2. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

```

Final cluster centroids:
Attribute                               Full Data      Cluster#
                                      (40.0)        0          1          2          3          4
                                      (40.0)        (6.0)      (19.0)     (8.0)     (3.0)     (4.0)
-----
edad                                   21            20         21         22         20         23
genero                                 0             0           0           0           0           0
sector_procedencia                     0             0           0           0           0           0
provincia_nacimiento                   11            19         11         11         7           11
trabaja                                 0             0           0           0           1           0
tipo_colegio                            2             2           2           2           2           2
CALCULO_INTEGRAL                       1             1           1           1           1           1
ARQUITECTURA_DE_ORDENADORES           1             1           1           1           1           1
ESTADISTICA_ANALITICA                   1             1           1           1           0           1
ESTRUCTURAS_DE_DATOS_Y_ALGORITMOS_FUNDAMENTALES 1             1           1           1           1           1
BANCO_DE_DATOS                          1             1           1           1           1           1
CATEDRA_INTEGRADORA:_PROYECTO_DE_VINCULACION:_CENTRO_DE_ASESORIA_TECNOLOGICA 1             1           1           1           1           1

```

Figura 278. Clústers obtenidos por Weka.



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

**Clúster\_0 (15%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Zamora Chinchipe, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Cálculo Integral, Arquitectura de Ordenadores, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales, Banco de Datos y Catedra Integradora; Proyecto de Vinculación “Centro de Asesoría Tecnológica”.

**Clúster\_1 (48%):** Estudiantes que tengan una edad de 21 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Cálculo Integral, Arquitectura de Ordenadores, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales, Banco de Datos y Catedra Integradora; Proyecto de Vinculación “Centro de Asesoría Tecnológica”.

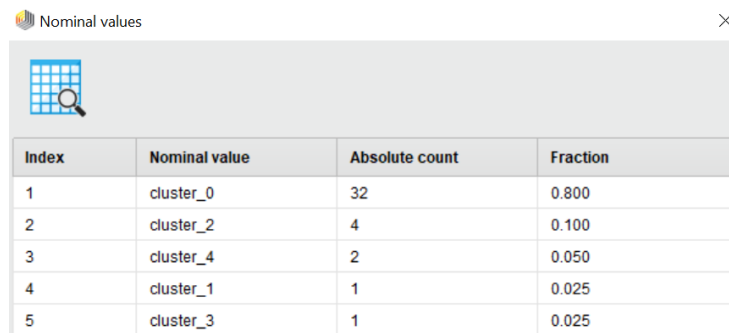
**Clúster\_2 (20%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Zamora Chinchipe, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Cálculo Integral, Arquitectura de Ordenadores, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales, Banco de Datos y Catedra Integradora; Proyecto de Vinculación “Centro de Asesoría Tecnológica”.

**Clúster\_3 (8%):** Estudiantes que tengan una edad de 20 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de El Oro, que si ejerzan alguna actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Cálculo Integral, Arquitectura de Ordenadores, Estructura de Datos y Algoritmos Fundamentales, Banco de Datos y Catedra Integradora; Proyecto de Vinculación “Centro de Asesoría Tecnológica”, con excepción de la asignatura de Estadística Analítica.

**Clúster\_4 (10%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que no cuenten con una actividad laboral y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al tercer ciclo, estas son: Cálculo Integral, Arquitectura de Ordenadores, Estadística Analítica, Estructura de Datos y Algoritmos Fundamentales, Banco de Datos y Catedra Integradora; Proyecto de Vinculación “Centro de Asesoría Tecnológica”.

### Resultados obtenidos por Algoritmo K - means – RapidMiner

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústers.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	32	0.800
2	cluster_2	4	0.100
3	cluster_4	2	0.050
4	cluster_1	1	0.025
5	cluster_3	1	0.025

Figura 279. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 40 instancias están clasificadas o agrupadas a su clúster más cercano.

Open in Turbo Prep Auto Model Filter (40 / 40 examples): all

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	CALCULO IN...	ARQUITECT...	EST
1	1	cluster_0	23	0	0	11	0	2	1	1	1
2	2	cluster_4	21	0	0	11	0	2	1	1	1
3	3	cluster_2	20	0	0	19	0	2	1	1	1
4	4	cluster_4	21	0	0	10	0	2	1	1	1
5	5	cluster_0	27	0	0	11	1	3	1	1	0
6	6	cluster_0	27	0	0	11	0	2	1	1	1
7	7	cluster_4	21	1	0	11	0	2	1	1	1
8	8	cluster_4	19	0	1	11	0	2	1	1	1
9	9	cluster_4	20	0	0	11	0	2	1	1	1
10	10	cluster_3	25	0	0	7	0	2	1	0	0
11	11	cluster_4	22	0	0	11	0	2	1	1	1
12	12	cluster_4	22	0	1	11	0	2	1	1	1
13	13	cluster_4	21	1	0	11	0	2	1	1	1

Figura 280. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 4 con más instancias agrupadas.

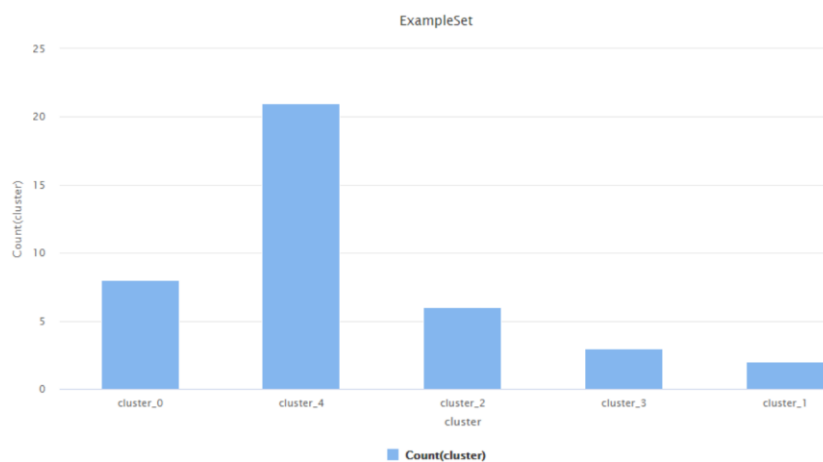


Figura 281. Representación gráfica de los clústers obtenidos

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide rojo y verde, se observa una relación entre los estudiantes de género masculino y femenino con una Institución Fiscomisional, mientras que en el centroide de color azul hace referencia a estudiantes de género Masculino con un colegio Fiscal.

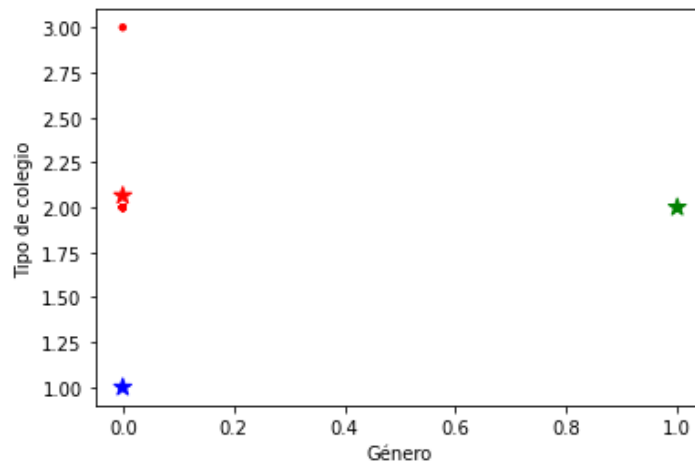


Figura 282. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Bancos de Datos** donde tiene una alta probabilidad de aprobar a dicha asignatura si su edad de 21, 24 y 34 años.

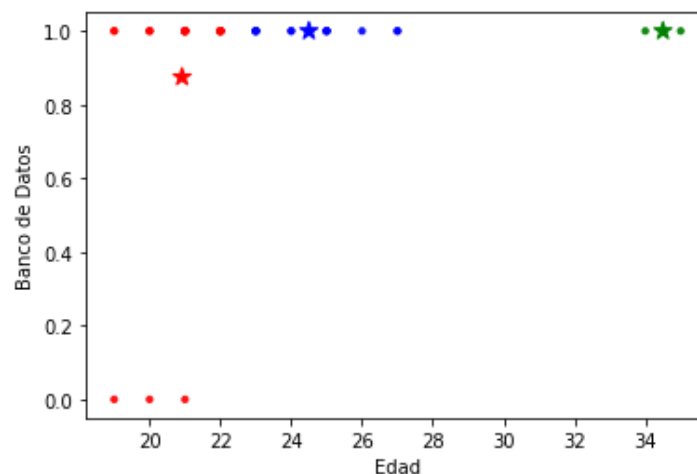


Figura 283. Relación Edad – Ingeniería de Software II

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** y **Bancos de Datos**, donde se observa que un grupo de estudiantes



pertencientes al género femenino tienen una alta probabilidad de aprobar a esta materia, por otro lado, los estudiantes de género masculino aprobarán sin inconvenientes a la misma, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprueban la asignatura de Bancos de Datos.

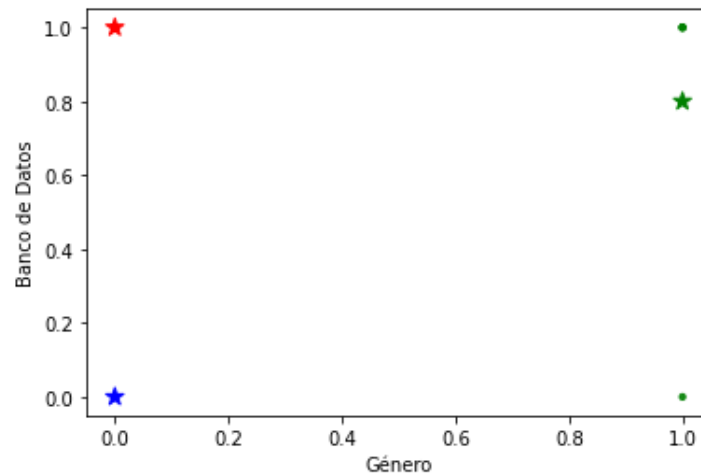


Figura 284. Relación Género – Ingeniería de Software II

Finalmente, la Figura 16 y 17 hacen referencia a la relación que se da entre materias, para la Fig. 16 se observa a las asignaturas de **Estructura de Datos y Algoritmos Fundamentales** y **Bancos de Datos**, en el clúster rojo se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, también se da el caso en donde apruebe la materia de Banco de Datos pero repruebe Estructura de Datos y Algoritmos Funcionales (clúster azul), finalmente se muestra una alta probabilidad que el estudiante no apruebe ninguna de las dos asignaturas (clúster verde).

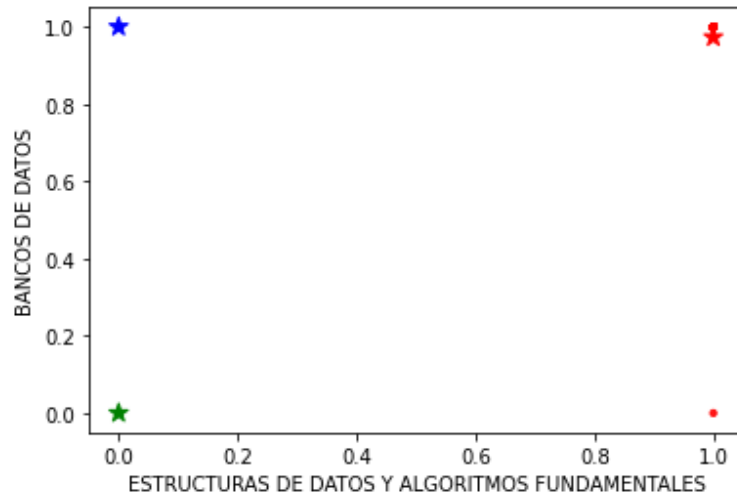
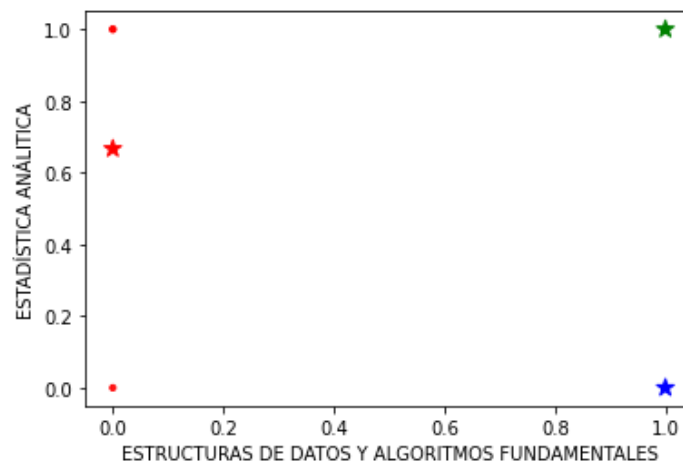


Figura 285. Relación Modelamiento Matemático – Ingeniería de Software II

Para los resultados obtenidos en la Fig. 17, se tiene a las asignaturas de Estadística Analítica y Estructura de Datos y Algoritmos Fundamentales, en donde, el clúster verde indica la aprobación en paralelo de ambas materias, el clúster azul hace referencia a estudiantes que únicamente aprueban la materia de Estructura de Datos y Algoritmos Fundamentales, por último, el clúster rojo hace hincapié a estudiantes con probabilidad media de aprobar la materia de Estadística Analítica y reprobado la otra asignatura.





## Anexo 17. Ingeniería en Computación – Cuarto ciclo

Dentro del proceso de Minería de datos Educativa aplicada al descubrimiento de patrones existentes en los registros académicos de los estudiantes de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja, se ha determinado trabajar de manera particular por cada ciclo de ambas carreras, dicho anexo especifica los modelos encontrados una vez aplicado cada algoritmo, correspondientes al **cuarto ciclo de la carrera de Ingeniería en Computación** para ello se ha dividido en dos secciones, la primera, algoritmo a priori, a través de las reglas de asociación; y la segunda clúster, con el algoritmo K-means, esto trabajado dentro de las herramientas Weka, RapidMiner y Python respectivamente.

### Resultados obtenidos por Algoritmo A priori – Weka

A través del análisis de correlación y con la decisión del investigador se ha determinado ciertas variables o características con las cuales se puede obtener el modelo para este algoritmo, en este caso, la **Figura 1**, muestra algunas de ellas y con las mismas se procede a seleccionar el algoritmo a priori, y como parámetro se realiza la creación de 200 reglas.

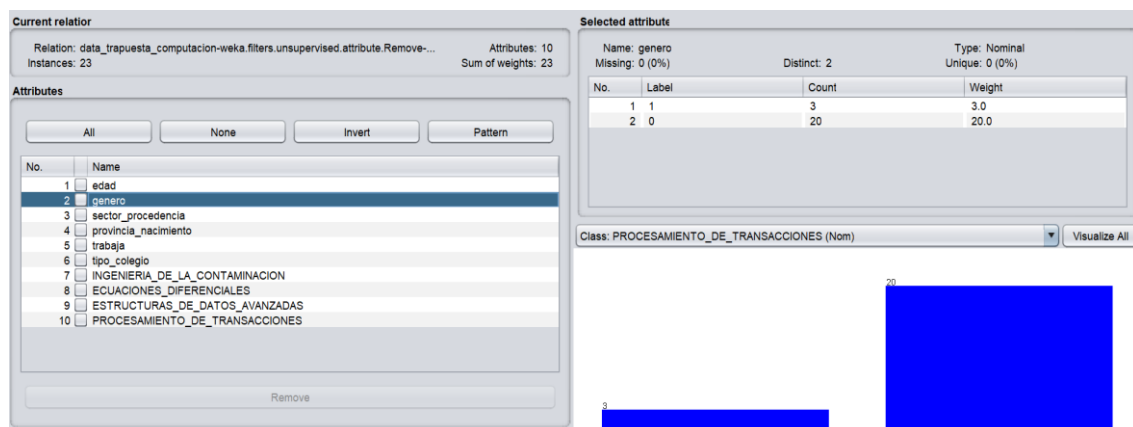
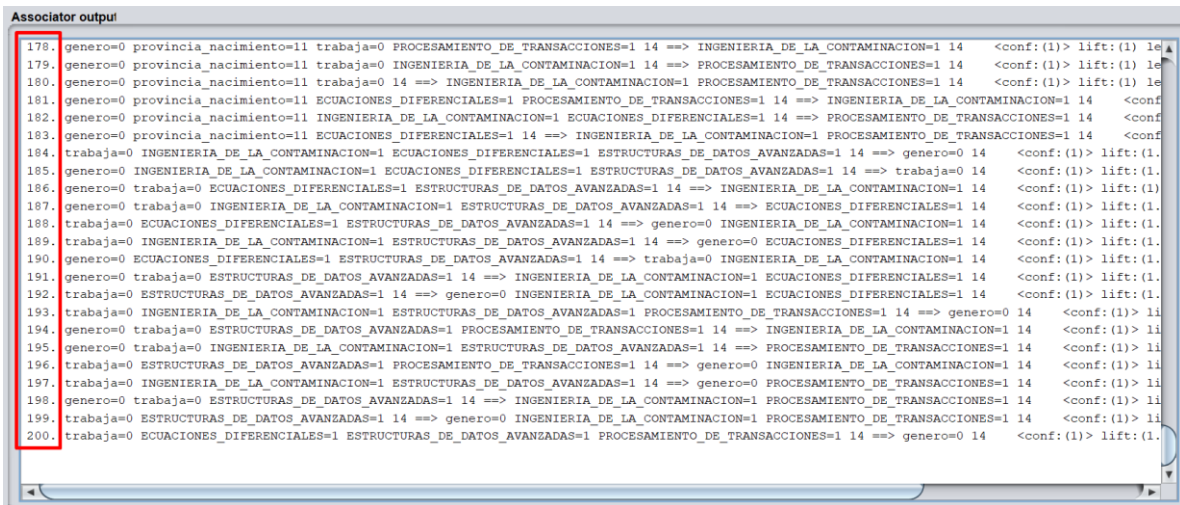


Figura 286. Selección de variables para la elaboración del modelo

Para la filtración y obtención de las reglas se hecho uso de las métricas lift y conf. Si al filtrar los datos se obtiene un número elevado de reglas se procederá a tomar 20 reglas como número máximo, caso contrario se trabajaría con las reglas obtenidas por los filtros aplicados.



```

178. genero=0 provincia_nacimiento=11 trabaja=0 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> lift:(1) le
179. genero=0 provincia_nacimiento=11 trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 14 ==> PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> lift:(1) le
180. genero=0 provincia_nacimiento=11 trabaja=0 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> lift:(1) le
181. genero=0 provincia_nacimiento=11 ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf
182. genero=0 provincia_nacimiento=11 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14 ==> PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf
183. genero=0 provincia_nacimiento=11 ECUACIONES_DIFERENCIALES=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf
184. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 14 <conf:(1)> lift:(1.
185. genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> trabaja=0 14 <conf:(1)> lift:(1.
186. genero=0 trabaja=0 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> lift:(1.
187. genero=0 trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 14 <conf:(1)> lift:(1.
188. trabaja=0 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> lift:(1.
189. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14 <conf:(1)> lift:(1.
190. genero=0 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> lift:(1.
191. genero=0 trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14 <conf:(1)> lift:(1.
192. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14 <conf:(1)> lift:(1.
193. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 14 <conf:(1)> li
194. genero=0 trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> li
195. genero=0 trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> li
196. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 14 <conf:(1)> li
197. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> li
198. genero=0 trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> li
199. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 <conf:(1)> li
200. trabaja=0 ECUACIONES_DIFERENCIALES=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 14 <conf:(1)> lift:(1.
  
```

Figura 287. Obtención de reglas de asociación

Una vez obtenidas las reglas (Ver Figura 2), para la selección de las reglas máximas, se realizó una copia de los resultados en un libro de Excel dentro del cual se procedió a separar el texto y a obtener únicamente las reglas con los parámetros previamente seleccionados.

Aplicado los filtros para reglas con nivel de confianza del 100% y lift mayor a 1.35 respectivamente, se obtiene un total de 11 reglas (Ver Figura 3).

REGLAS	CONF	LIFT	LEV	CONV
135. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
148. genero=0 trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
149. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.44)	lev:(0.19)	[4] conv:(4.26)
167. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
168. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
172. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
174. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
187. genero=0 trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
189. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.44)	lev:(0.19)	[4] conv:(4.26)
191. genero=0 trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.35)	lev:(0.16)	[3] conv:(3.65)
192. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14	conf:(1)>	lift:(1.44)	lev:(0.19)	[4] conv:(4.26)

Figura 288. Reglas obtenidas aplicando filtros (Conf – Lift)

A continuación, se describen las premisas obtenidas:

- 1) Si un estudiante no ejerce ninguna actividad laboral y aprueba la Estructura de Datos Avanzadas es muy probable que también apruebe la materia de Ecuaciones Diferenciales.
- 2) Si un estudiante es de género masculino, no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que también apruebe la materia de Ecuaciones Diferenciales.



- 3) Si un estudiante no ejerce actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que sea de género masculino y que apruebe la materia de Ecuaciones Diferenciales.
- 4) Si un estudiante no ejerce actividad laboral y aprueba la materia de Ingeniería de la Contaminación, es muy probable que sea de género masculino y que apruebe la materia de Ecuaciones Diferenciales.
- 5) Si un Estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que apruebe las materias de Ingeniería de Computación y Ecuaciones Diferenciales.
- 6) Si un Estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada y Procesamiento de Transacciones, es muy probable que apruebe la materia de Ecuaciones Diferenciales.
- 7) Si un estudiante no ejerce ninguna actividad laboral y aprueba las materias de Estructura de Datos Avanzada y Procesamiento de Transacciones, es muy probable que también apruebe la materia de Ecuaciones Diferenciales.
- 8) Si un estudiante es de género masculino, no ejerce ninguna actividad laboral y aprueba las materias de Ingeniería en Contaminación y Estructura de Datos Avanzada, es muy probable que también apruebe la materia de Ecuaciones Diferenciales.
- 9) Si un estudiante no ejerce ninguna actividad laboral y aprueba las materias de Ingeniería de la Contaminación y Estructura de Datos Avanzada, es muy probable que sea de género masculino y apruebe la materia de Ecuaciones Diferenciales.
- 10) Si un estudiante de género masculino, no trabaja y aprueba la materia de Estructura de Datos Avanzada, es muy probable que también apruebe las materias de Ingeniería de la Contaminación y Ecuaciones Diferenciales.
- 11) Si un estudiante no trabaja y aprueba la materia de Estructura de Datos Avanzada, es muy probable que también apruebe las materias de Ingeniería de la Contaminación y Ecuaciones Diferenciales.

Por otra parte, se han determinado Reglas definiendo lift como predeterminado, obteniendo valores del 1.44 y una confiabilidad de las premisas mayores a 88% (Ver Figura 4) con lo que se obtienen los siguientes resultados:

REGLAS	CONF	LIFT	LEV	COV	CONV
1. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
2. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
3. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
4. genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
5. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
6. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
7. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
8. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
9. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
10. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
11. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
12. genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
13. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
14. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
15. genero=0 ECUACIONES_DIFERENCIALES=1 16 ==> trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
16. genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 PROCESAMIENTO_DE_TRANSACCIONES=1 16 ==> trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 14 con	:{0.88}	lift:(1.44)>	lev:(0.19)	[4]	conv:(2.09)
17. trabaja=0 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 INGENIERIA_DE_LA_CONTAMINACION=1 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)
18. trabaja=0 INGENIERIA_DE_LA_CONTAMINACION=1 ESTRUCTURAS_DE_DATOS_AVANZADAS=1 PROCESAMIENTO_DE_TRANSACCIONES=1 14 ==> genero=0 ECUACIONES_DIFERENCIALES=1 14 con	:{1}	lift:(1.44)>	lev:(0.19)	[4]	conv:(4.26)

Figura 289. Reglas obtenidas en Weka con Lift mayor a 1

- 1) Si un estudiante es de género masculino y aprueba las materias de Ecuaciones diferenciales, es muy probable que no ejerza alguna actividad laboral y que apruebe la materia de Estructura de Datos Avanzada.
- 2) Si un estudiante no cuenta con alguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que sea de género masculino y además apruebe la materia de Ecuaciones Diferenciales.
- 3) Si un estudiante es de género masculino y aprueba las materias de Ecuaciones diferenciales, es muy probable que no ejerza alguna actividad laboral y que apruebe las materias de Ingeniería de la Contaminación y Estructura de Datos Avanzada.
- 4) Si un estudiante es de género masculino y aprueba las materias de Ingeniería de la Contaminación y Ecuaciones diferenciales, es muy probable que no ejerza alguna actividad laboral y la materia de Estructura de Datos Avanzada.
- 5) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que sea de género masculino y apruebe la materia de Ingeniería de la Contaminación y Ecuaciones Diferenciales.
- 6) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Ingeniería de la Contaminación y Estructura de Datos Avanzada, es muy probable que sea de género masculino y apruebe la materia de Ecuaciones Diferenciales.



- 7) Si un estudiante es de género masculino y aprueba la materia de Ecuaciones Diferenciales, es muy probable que no realice ninguna actividad laboral y apruebe las asignaturas de Estructura de Datos Avanzada y Procesamiento de Transacciones.
- 8) Si un estudiante no ejerce actividad laboral y aprueba la asignatura de Estructura de Datos Avanzada, es muy probable que sea de género masculino y que apruebe también las materias de Ecuaciones Diferenciales y Procesamiento de Transacciones.
- 9) Si un estudiante es de género masculino y aprueba las materias de Ecuaciones Diferenciales y Procesamiento de Transacciones, es muy probable que no ejerza actividad laboral y apruebe la asignatura de Estructura de Datos Avanzada.
- 10) Si un estudiante no realiza ninguna actividad laboral y aprueba las asignaturas de Estructura de Datos Avanzada y Procesamiento de Transacciones, es muy probable que sea de género masculino y apruebe la materia de Ecuaciones Diferenciales.
- 11) Si un estudiante de género masculino aprueba la materia de Ecuaciones Diferenciales, es muy probable que no ejerza ninguna actividad laboral y apruebe las materias de Ingeniería de la Comunicación, Estructura de Datos Avanzada y Procesamiento de Transacciones.
- 12) Si un estudiante de género masculino aprueba la materia de Ingeniería de la Contaminación y Ecuaciones Diferenciales, es muy probable que no ejerza ninguna actividad laboral y también apruebe la asignatura de Estructura de Datos Avanzada y Procesamiento de Transacciones.
- 13) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que sea de género masculino y que apruebe las asignaturas de Ingeniería de la Contaminación, Ecuaciones Diferenciales y Procesamiento de Transacciones.
- 14) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Ingeniería de la Contaminación y Estructura de Datos Avanzada, es muy probable que sea de género masculino y que apruebe las asignaturas de Ingeniería de la Contaminación, Ecuaciones Diferenciales y Procesamiento de Transacciones.





- 15) Si un estudiante de género masculino aprueba la materia de Ecuaciones Diferenciales y Procesamiento de Transacciones, es muy probable que no ejerza ninguna actividad laboral y también apruebe la asignatura de Ingeniería de la Contaminación y Estructura de Datos Avanzada.
- 16) Si un estudiante de género masculino aprueba la materia de Ingeniería de la Contaminación y Ecuaciones Diferenciales y Procesamiento de Transacciones, es muy probable que no ejerza ninguna actividad laboral y también apruebe la asignatura de Estructura de Datos Avanzada.
- 17) Si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada y Procesamiento de Transacciones, es muy probable que sea de género masculino y que apruebe las asignaturas de Ingeniería de la Contaminación, Ecuaciones Diferenciales.
- 18) Si un estudiante no trabaja, aprueba la materia de Ingeniería de la Contaminación y Estructura de Datos Avanzada y Procesamiento de Transacciones, es muy probable que sea de género masculino y que también apruebe la asignatura de Estructura de Datos Avanzada.

En este caso al tomar al lift como métrica principal, se obtiene reglas con resultados mejor predictivos que al azar o de manera random. En este caso se ha identificado un número de reglas considerables con un lift elevado, mismas que dan mayor predicción en un futuro para este curso en estudiantes inferiores, el valor esta métrica (1.44) se enlaza con la regla obtenida, en este caso, en ciertas reglas, se identifica el 100% de veracidad en las mismas. Sin embargo, por ejemplo las regla # 1 , hace referencia a que los sí un estudiante es de género masculino y aprueba las materias de Ecuaciones diferenciales, es muy probable que no ejerza alguna actividad laboral y que apruebe la materia de Estructura de Datos Avanzado; la regla # 3, muestra una estrecha relación, esta nos indica que si un estudiante es de género masculino y aprueba las materias de Ecuaciones diferenciales, es muy probable que no ejerza alguna actividad laboral y que apruebe las materias de Ingeniería de la Contaminación y Estructura de Datos Avanzada, y algunas otras reglas que se asemejan a las descritas con anterioridad, para ellas el nivel de confianza es del 88%, que si bien es cierto no es “malo” pero en relación a las reglas con el 100% se podrían descartar.



La forma paralela para aprobar ciertas asignaturas en relación a otras se sigue identificando, como el caso de la regla # 5 indica que, si un estudiante no ejerce ninguna actividad laboral y aprueba la materia de Estructura de Datos Avanzada, es muy probable que sea de género masculino y apruebe la materia de Ingeniería de la Contaminación y Ecuaciones Diferenciales. De la misma manera se sigue manteniendo la ausencia del género femenino a la altura de dichos cursos es baja.

### **Resultados obtenidos por Algoritmo A priori – RapidMiner**

Las reglas obtenidas a través de esta herramienta son muy similares a los resultados obtenidos en Weka. La relación entre las asignaturas se sigue manteniendo, se muestra la presencia de estudiante provenientes de la Provincia de Loja, además, en este caso se muestra la característica tipo de colegio del cual proviene el estudiante, si bien es cierto, las bases que los estudiantes reciben en las diferentes instituciones educativas ya sean Fiscales, Fiscomisionales y Privadas influyen mucho en el desempeño de los estudiantes al momento de ingresar a una carrera universitaria, puesto que, a partir de estas bases se pueden formar y obtener nuevos conocimientos. Otra de las variables presentes en este conjunto de premisas es la edad y el género; la edad como son primero ciclos la edad varía entre 24 a 32 años, la mayoría de edad se debe a la homologación de ciertos estudiantes, y de la misma manera se refleja la persistencia del género masculino. Esto se ve reflejado en la Figura 5.

## AssociationRules

### Association Rules

```
[ECUACIONES DIFERENCIALES] --> [PROCESAMIENTO DE TRANSACCIONES, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[PROCESAMIENTO DE TRANSACCIONES, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [provincia_nacimiento, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[provincia_nacimiento, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [tipo_colegio, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[tipo_colegio, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [estado_civil, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[estado_civil, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, PROCESAMIENTO DE TRANSACCIONES, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, ECUACIONES DIFERENCIALES] --> [PROCESAMIENTO DE TRANSACCIONES, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[PROCESAMIENTO DE TRANSACCIONES, ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, PROCESAMIENTO DE TRANSACCIONES, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, provincia_nacimiento, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, ECUACIONES DIFERENCIALES] --> [provincia_nacimiento, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[provincia_nacimiento, ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, provincia_nacimiento, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, tipo_colegio, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, ECUACIONES DIFERENCIALES] --> [tipo_colegio, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[tipo_colegio, ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, tipo_colegio, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, estado_civil, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, ECUACIONES DIFERENCIALES] --> [estado_civil, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[estado_civil, ECUACIONES DIFERENCIALES] --> [INGENIERIA DE LA CONTAMINACION, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[INGENIERIA DE LA CONTAMINACION, estado_civil, ECUACIONES DIFERENCIALES] --> [ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[ECUACIONES DIFERENCIALES] --> [PROCESAMIENTO DE TRANSACCIONES, provincia_nacimiento, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[PROCESAMIENTO DE TRANSACCIONES, ECUACIONES DIFERENCIALES] --> [provincia_nacimiento, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
[provincia_nacimiento, ECUACIONES DIFERENCIALES] --> [PROCESAMIENTO DE TRANSACCIONES, ESTRUCTURAS DE DATOS AVANZADAS] (confidence: 0.882)
```

Figura 290. Reglas de asociación en RapidMiner

## Resultados obtenidos por Algoritmo A priori – Python

Al ejecutar este algoritmo, se ha podido obtener ciertas premisas que cumplen con las métricas descritas en un inicio (Ver Figuras 3 y 4), como resultado observado en la Figura 6, se logra identificar que, si bien es cierto ya se muestra más asignaturas en comparación a ciclos inferiores, pero en este caso la relación entre las asignaturas relacionadas con las matemáticas con estudiantes de género masculino es más visibles, a esto le acompaña algunas otras características como no ejercer ninguna actividad laboral, ser solteros y ser oriundos de la provincia de Loja. Estas premisas cuentan con una confiabilidad de hasta el 100% con un lift de 1.04, estas son las características más relevantes que se destacan dentro de las premisas.

```
RelationRecord(items=frozenset({'privado', 'PC_A', 'NTrabaja'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
RelationRecord(items=frozenset({'Urbano', 'Soltero', 'PC_A'}), support=0.82680869565217391, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'})
RelationRecord(items=frozenset({'IC_A', 'PC_A', '11', 'Soltero'}), support=0.782608695652174, ordered_statistics=[OrderedStatistic(items_base=frozenset({'11'})
RelationRecord(items=frozenset({'IC_A', 'Soltero', 'PC_A', 'ED_A'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'ED_A'})
RelationRecord(items=frozenset({'NTrabaja', 'IC_A', 'PC_A', 'Masculino'}), support=0.782608695652174, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
RelationRecord(items=frozenset({'NTrabaja', 'IC_A', 'Soltero', 'Masculino'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
RelationRecord(items=frozenset({'privado', 'IC_A', 'PC_A', 'NTrabaja'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
RelationRecord(items=frozenset({'Urbano', 'IC_A', 'Soltero', 'PC_A'}), support=0.82680869565217391, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Soltero'})
RelationRecord(items=frozenset({'NTrabaja', 'Soltero', 'PC_A', 'Masculino'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
RelationRecord(items=frozenset({'PC_A', 'Masculino', 'Soltero', 'IC_A', 'NTrabaja'}), support=0.7391304347826086, ordered_statistics=[OrderedStatistic(items_base=frozenset({'NTrabaja'})
```

Figura 291. Reglas de asociación obtenidas por Python



## Resultados obtenidos por Algoritmo K - means – Weka

Para los clústeres obtenidos en weka, haciendo uso del algoritmo K – means se ha realizado un limitante de 5 clúster, cada uno de ellos con las instancias que les corresponden, Ver Figura 6.

### Clustered Instances

0	1 ( 4%)
1	5 ( 22%)
2	11 ( 48%)
3	5 ( 22%)
4	1 ( 4%)

Figura 292. Clústers con sus respectivas instancias

Dentro de estos, los más representativos son el clúster 1, 2 y 3. La Figura 7, indica las características que componen a cada clúster, seguidamente se describe cada clúster:

### Final cluster centroids:

Attribute	Full Data (23.0)	Cluster#				
		0 (1.0)	1 (5.0)	2 (11.0)	3 (5.0)	4 (1.0)
edad	21	35	23	21	22	23
genero	0	0	0	0	0	0
sector_procedencia	0	0	0	0	0	0
provincia_nacimiento	11	11	11	11	11	11
trabaja	0	1	0	0	0	0
tipo_colegio	2	1	2	2	2	2
INGENIERIA_DE_LA_CONTAMINACION	1	1	1	1	1	1
ECUACIONES_DIFERENCIALES	1	1	0	1	1	1
ESTRUCTURAS_DE_DATOS_AVANZADAS	1	0	0	1	1	0
PROCESAMIENTO_DE_TRANSACCIONES	1	1	1	1	1	1

Figura 293. Clústers obtenidos por Weka.

**Clúster\_0 (44%):** Estudiantes que tengan una edad de 35 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que si ejerza alguna actividad laboral a parte de sus académicas y que hayan finalizado sus estudios secundarios en una institución Fiscal, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, estas son: Ingeniería de la Contaminación, Ecuaciones Diferenciales y Procesamiento de Transacciones, excepto Estructura de Datos Avanzadas.

**Clúster\_1 (24%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja,



UNL

Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

que no ejerza alguna actividad laboral a parte de sus académicas y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán ciertas materias relacionadas a las matemáticas correspondientes al cuarto ciclo, estas son: Ingeniería de la Contaminación y Procesamiento de Transacciones pero reprobarán a Ecuaciones Diferenciales, Estructura de Datos Avanzadas

**Clúster\_2 (23%):** Estudiantes que tengan una edad de 21 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que si ejerza alguna actividad laboral a parte de sus académicas y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, estas son: Ingeniería de la Contaminación, Ecuaciones Diferenciales, Estructura de Datos Avanzadas y Procesamiento de Transacciones.

**Clúster\_3 (4%):** Estudiantes que tengan una edad de 22 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que si ejerza alguna actividad laboral a parte de sus académicas y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, estas son: Ingeniería de la Contaminación, Ecuaciones Diferenciales, Estructura de Datos Avanzadas y Procesamiento de Transacciones.

**Clúster\_4 (6%):** Estudiantes que tengan una edad de 23 años, pertenezcan al género masculino, sean del sector Urbano y provenientes de la provincia de Loja, que si ejerza alguna actividad laboral a parte de sus académicas y que hayan finalizado sus estudios secundarios en una institución Fiscomisional, aprobarán todas las materias relacionadas a las matemáticas correspondientes al cuarto ciclo, estas son: Ingeniería de la Contaminación, Ecuaciones Diferenciales y Procesamiento de Transacciones, a excepción de Estructura de Datos Avanzadas.

### **Resultados obtenidos por Algoritmo K - means – RapidMiner**

La Figuras 9 se muestra la agrupación de cada clúster con sus respectivas características, además presenta las instancias correspondientes a cada uno de ellos. De la misma manera en RapidMiner, al igual que Weka se ha obtenido un total de 5 clústers.

Nominal values

Index	Nominal value	Absolute count	Fraction
1	cluster_4	10	0.435
2	cluster_2	6	0.261
3	cluster_0	3	0.130
4	cluster_1	2	0.087
5	cluster_3	2	0.087

Figura 294. Clúster con total de instancias

Por otra parte, en la Figura 10, se identifica cada una de las características correspondientes a cada clúster, es decir, cada una de las 23 instancias están clasificadas o agrupadas a su clúster más cercano.

Open in Turbo Prep Auto Model Filter (23 / 23 examples): all

Row No.	id	cluster	edad	genero	sector_proc...	provincia_n...	trabaja	tipo_colegio	INGENIERIA ...	ECUACIONE...	EST
1	1	cluster_4	23	0	0	11	0	2	1	0	0
2	2	cluster_2	21	0	0	11	0	2	1	1	1
3	3	cluster_2	21	1	0	11	0	2	1	0	0
4	4	cluster_2	19	0	1	11	0	2	1	1	1
5	5	cluster_3	25	0	0	7	0	2	1	1	1
6	6	cluster_4	22	0	0	11	0	2	1	1	1
7	7	cluster_4	22	0	1	11	0	2	1	1	1
8	8	cluster_4	24	0	0	11	0	3	1	1	1
9	9	cluster_4	22	0	0	11	1	2	1	0	1
10	10	cluster_2	21	0	0	11	0	1	1	1	1
11	11	cluster_2	21	0	0	11	0	2	1	1	1
12	12	cluster_0	26	0	1	19	0	2	1	0	0
13	13	cluster_4	22	0	0	11	0	2	1	1	1

Figura 295. Agrupación de cada Instancia a su respectivo Clúster

De igual forma, en la Figura 11 se identifica de manera gráfica como se encuentra cada clúster con el valor aproximado de instancias pertenecientes a los mismos, siendo el clúster 4 con más instancias agrupadas.

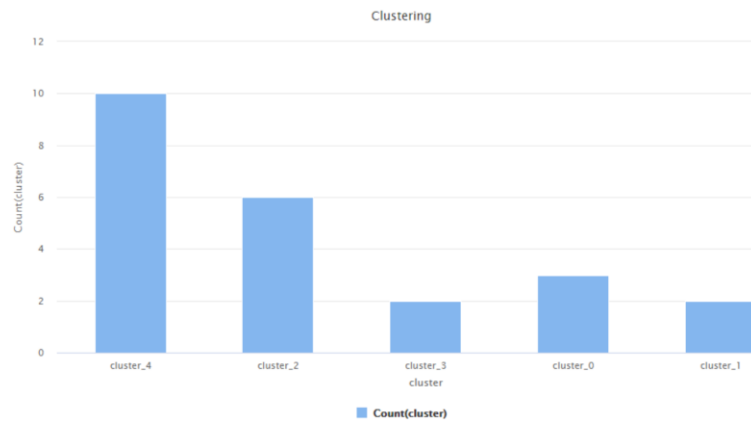


Figura 296. Representación gráfica de los clústeres obtenidos

### Resultados obtenidos por Algoritmo K - means – Python

Para la obtención de los resultados a través de este algoritmo, se ha realizado un análisis acerca de cuantos clústers son los necesarios de obtener (valor de K), esto se demuestra a través del “Método de codo” o “punto de codo” ejecutado dentro de esta herramienta, el cual arroja como resultado un número adecuado de 3 clústers. De la misma manera, en la representación gráfica se puede realizar la relación únicamente entre 2 y 3 variables, en este caso se ha decidido trabajar con dos variables.

Dentro de la gráfica 13, se han obtenido los clústers con las variables **Género** y **Tipo de colegio**, en los cuales se puede observar a cada uno de los clústers con sus respectivos centroides; considerado al centroide rojo y verde, se observa una relación entre los estudiantes de género masculino y femenino con una Institución Fiscomisional, mientras que en el centroide de color azul hace referencia a estudiantes de género masculino con un colegio Fiscal, dentro de esta representación no se identifica un colegio Particular, esta características ha prevalecido en ciclos inferiores, donde se identifica que los estuantes que prevalecen son más de una Institución Fiscomisional.

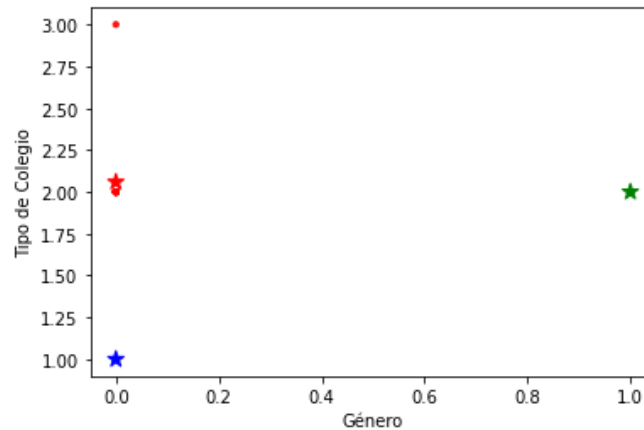


Figura 297. Relación Género – Tipo de Colegio

En la figura 14, se muestra la relación entre la **edad** y la asignatura **Ecuaciones Diferenciales** en donde se tiene una alta probabilidad de aprobar a dicha asignatura si los estudiantes tienen una edad de 22, 24 y 34 años.

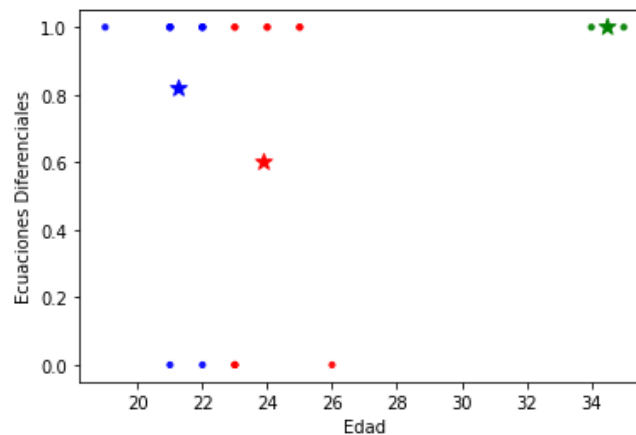


Figura 298. Relación Edad – Ingeniería de Software II

Por otra parte, dentro de la Figura 15 se muestra la relación entre el **Género** e **Ecuaciones Diferenciales**, donde se observa que un grupo de estudiantes pertenecientes al género femenino tienen una baja probabilidad de aprobar la materia. Por otro lado, el género masculino tiene mucha más probabilidad de aprobarla, sin embargo, hay un número de estudiantes pertenecientes al mismo género que reprobaban la asignatura de Ecuaciones Diferenciales.

Este mismo comportamiento se refleja con la asignatura de Estructuras de Datos Avanzadas

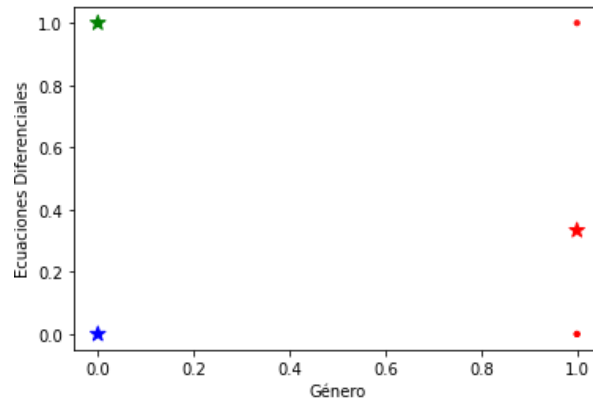


Figura 299. Relación Género – Ingeniería de Software II

Finalmente, la Figura 16 hace referencia a la relación que se da entre materias, en este caso **Estructura de Datos Avanzada** y **Ecuaciones Diferenciales**, en el clúster verde se refleja la aprobación de ambas materias, es decir, si aprueba una lo hará también en la otra, sin embargo, también se da el caso en donde los estudiantes tendrán una baja probabilidad de aprobar Estructuras de Datos Avanzadas, pero repruebe Ecuaciones Diferenciales (clúster rojo), finalmente se muestra una alta probabilidad que el estudiante únicamente apruebe Ecuaciones Diferenciales (clúster azul).

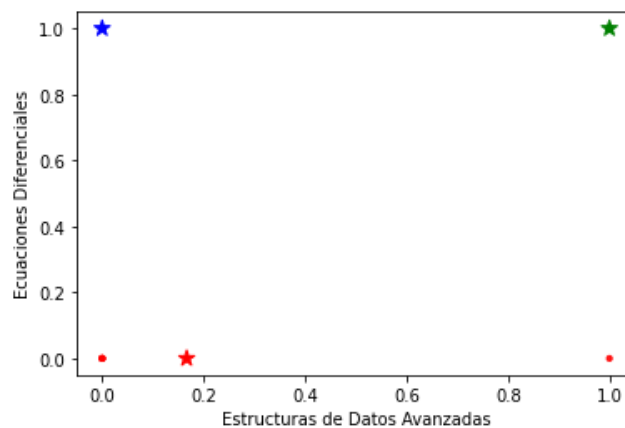


Figura 300. Relación Modelamiento Matemático – Ingeniería de Software II



UNL

Universidad Nacional de Loja



Carrera de Ingeniería en Sistemas / Computación

## Anexo 18. Certificación: Máster Universitario en Ingeniería Matemática y Computación



UNIVERSIDAD NACIONAL DE LOJA

FACULTAD DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES  
Carrera de Ingeniería en Sistemas / Carrera Computación



A quien corresponda

Certifico a través del presente documento, haber asesorado y validado la propuesta del "Plan de acción para mitigar la deserción Universitaria" planteado por la estudiante **Jhulissa Isabel Villamagua Poma** con CC: 1150692414 el mismo que forma parte del desarrollo del trabajo de titulación denominado "**MINERÍA DE DATOS EDUCATIVA APLICADA AL DESCUBRIMIENTO DE PATRONES EN REGISTROS ACADÉMICOS DE ESTUDIANTES DE INGENIERÍA CON ASIGNATURAS RELACIONADAS A LAS MATEMÁTICAS**".

La interesada puede hacer uso del presente certificado como considere necesario.

Loja 16 de agosto de 2021



Firmado electrónicamente por:  
ANDRES ROBERTO  
NAVAS CASTELLANOS

Ing. Andrés Roberto Navas Castellanos Mg.Sc  
Docente CIS-UNL  
[andres.navas@unl.edu.ec](mailto:andres.navas@unl.edu.ec)



Universidad  
Nacional  
de Loja



Carrera de Ingeniería en  
Sistemas / Computación

## Anexo 19. Certificación: Miembro del Consejo Consultivo - UNL



Universidad  
Nacional  
de Loja

FACULTAD DE LA ENERGÍA, LAS INDUSTRIAS Y  
LOS RECURSOS NATURALES NO RENOVABLES  
Carrera de Ingeniería en Sistemas / Computación

CERTIFICACIÓN Nro.: 2021-002-CZW-CIS-C-FEIRNNR-UNL-CERT

Loja, 18 de agosto de 2021.

Ing. Wilman Patricio Chamba Zaragocín  
Docente CIS-C - UNL

### CERTIFICO:

Haber asesorado y validado la Propuesta del Plan de acción para mitigar la deserción Universitaria planteado por la estudiante **Jhulissa Isabel Villamagua Poma** con CI: **1150692414**, el mismo que forma parte del desarrollo del Trabajo de Titulación denominado **“MINERÍA DE DATOS EDUCATIVA APLICADA AL DESCUBRIMIENTO DE PATRONES EN REGISTROS ACADÉMICOS DE ESTUDIANTES DE INGENIERÍA CON ASIGNATURAS RELACIONADAS A LAS MATEMÁTICAS”**.

Es todo lo que puedo certificar en honor a la verdad.



Firmado electrónicamente por:  
WILMAN PATRICIO  
CHAMBA ZARAGOCIN

Wilman Chamba Zaragocín  
DOCENTE CIS-UNL  
[wpchamba@unl.edu.ec](mailto:wpchamba@unl.edu.ec)





UNL

Universidad Nacional de Loja



Carrera de Ingeniería en Sistemas / Computación

## Anexo 20. Certificación: Magíster en Docencia Universitaria e Investigación Educativa



UNL

Universidad Nacional de Loja

FACULTAD DE LA EDUCACIÓN, EL ARTE,  
Y LA COMUNICACIÓN  
CARRERA DE PEDAGOGÍA DE LOS IDIOMAS  
NACIONALES Y EXTRANJEROS

Loja, 25 de Agosto de 2021

Dra. Marcia Iliana Criollo Vargas

DOCENTE DE LA CARRERA DE PEDAGOGÍA DE LOS IDIOMAS NACIONALES Y EXTRANJEROS -  
UNL

### CERTIFICA:

Haber revisado y validado la Propuesta del Plan de Acción para mitigar la deserción Universitaria planteado por la estudiante Jhulissa Isabel Villamagua Poma con CI: 1150692414, el mismo que forma parte del Trabajo de Titulación denominado "MINERÍA DE DATOS EDUCATIVA APLICADA AL DESCUBRIMIENTO DE PATRONES EN REGISTROS ACADÉMICOS DE ESTUDIANTES DE INGENIERÍA CON ASIGNATURAS RELACIONADA A LAS MATEMÁTICAS".

Es cuanto puede certificar en honor a la verdad, autorizando a la peticionaria hacer uso del presente para fines legales.

Lo certifica:



El código de verificación es:  
MARCIA ILIANA  
CRIOLLO  
VARGAS

Dra. Marcia Criollo Vargas Mg. Sc.  
DOCENTES CARRERA PINE-UNL  
marcia.criollo@unl.edu.ec



UNL

Universidad  
Nacional  
de Loja

Facultad de la Energía, las Industrias y los Recursos Naturales no Renovables

---

CARRERA DE INGENIERÍA EN SISTEMAS

# Minería de Datos Educativa aplicada al descubrimiento de patrones en registros académicos de estudiantes de Ingeniería con asignaturas relacionadas a las **matemáticas**

Línea de investigación: Sistemas Inteligentes

PROYECTO DE TITULACIÓN EN  
INGENIERÍA EN SISTEMAS.

**Autor:**

- ◇ Jhulissa Isabel Villamagua Poma, jivillamaguap@unl.edu.ec

**Tutor:**

- Luis Antonio Chamba-Eras, Mg.Sc.



Carrera de Ingeniería en  
Sistemas / Computación

LOJA - ECUADOR

2020

## Certificación del Tutor

**Luis Antonio Chamba-Eras, Mg.Sc.** en calidad de tutor voluntario del proyecto de trabajo de titulación, certifico la tutela a Jhulissa Isabel Villamagua Poma, con el tema **Minería de Datos Educativa aplicada al descubrimiento de patrones en registros académicos de estudiantes de Ingeniería con asignaturas relacionadas a las matemáticas**, quien ha cumplido con todas las observaciones requeridas. Es todo cuanto puedo decir en honor a la verdad, facultando al interesado hacer uso de la presente, así como el trámite de pertinencia del presente proyecto.

Loja, 5 de marzo de 2020

Atentamente,  
Luis Antonio Chamba-Eras, Mg.Sc.  
**TUTOR**

---

## Certificación de Autoría del Proyecto

Yo Jhulissa Isabel Villamagua Poma, estudiante de la Universidad Nacional de Loja, declaro en forma libre y voluntaria que el presente proyecto de trabajo de titulación que versa sobre **Minería de Datos Educativa aplicada al descubrimiento de patrones en registros académicos de estudiantes de Ingeniería con asignaturas relacionadas a las matemáticas**, así como la expresiones vertidas en la misma son autoría del compareciente, quien ha realizado en base a recopilación bibliográfica primaria y secundaria. En consecuencia asumo la responsabilidad de la originalidad de la misma y el cuidado al remitirse a las fuentes bibliográficas respectivas para fundamentar el contenido expuesto.

Atentamente,  
Jhulissa Isabel Villamagua Poma

---

# Índice general

<b>1. Problemática</b>	<b>1</b>
1.1. Situación Problemática . . . . .	1
1.2. Problema de Investigación . . . . .	3
<b>2. Justificación</b>	<b>4</b>
<b>3. Objetivos</b>	<b>6</b>
3.1. General . . . . .	6
3.2. Específicos . . . . .	6
<b>4. Alcance</b>	<b>7</b>
<b>5. Marco Teórico</b>	<b>8</b>
5.1. Deserción Estudiantil . . . . .	8
5.1.1. Deserción Universitaria en Ecuador . . . . .	8
5.1.2. Factores que causan la deserción Universitaria . . . . .	8
5.2. Minería de Datos . . . . .	10
5.2.1. Minería de Datos Aplicada al campo Educativo . . . . .	10
5.3. Trabajos Relacionados . . . . .	11
<b>6. Metodología</b>	<b>25</b>
<b>7. Cronograma</b>	<b>26</b>
<b>8. Presupuesto</b>	<b>28</b>
<b>Bibliografía</b>	<b>32</b>
<b>Lista de Acrónimos y Abreviaturas</b>	<b>36</b>
<b>A. Anexo I</b>	<b>37</b>
A.1. Entrevistas a Estudiantes y Docentes CIS-CC . . . . .	37
<b>B. Anexo II</b>	<b>39</b>
B.1. Revisión Sistemática de Literatura . . . . .	39

# Índice de tablas

5.1. Trabajos Relacionados	12
5.1. Trabajos Relacionados	13
5.1. Trabajos Relacionados	14
5.1. Trabajos Relacionados	15
5.1. Trabajos Relacionados	16
5.1. Trabajos Relacionados	17
5.1. Trabajos Relacionados	18
5.1. Trabajos Relacionados	19
5.1. Trabajos Relacionados	20
5.1. Trabajos Relacionados	21
5.1. Trabajos Relacionados	22
5.1. Trabajos Relacionados	23
5.1. Trabajos Relacionados	24
7.1. Etiquetas	26
8.1. Presupuesto de Talento Humano	28
8.2. Presupuesto de Software y Hardware	29
8.3. Servicios	30
8.4. Presupuesto General	31
B.1. Cadenas de Búsqueda	40
B.2. Trabajos Relacionados	40
B.2. Trabajos Relacionados	41
B.2. Trabajos Relacionados	42

**Minería de Datos Educativa  
aplicada al descubrimiento de  
patrones en registros académicos de  
estudiantes de Ingeniería con  
asignaturas relacionadas a las  
matemáticas**

Línea de investigación: Sistemas Inteligentes





# 1. Problemática

## 1.1. Situación Problemática

Según datos presentados por el Instituto Nacional de Evaluación Educativa (INEVAL) de Ecuador, en el programa para la Evaluación de Estudiantes para el desarrollo (PISA-D) desarrollada en el año 2017, el nivel mínimo de habilidades en Matemática alcanza únicamente el 29% reflejándose así una alta deficiencia en la formación de los estudiantes desde etapas iniciales [1]. De igual forma, en los resultados de los exámenes obtenidos por estudiantes de secundaria denominado “Ser Bachiller”, en el año 2017 - 2018, se ha demostrado que el grado de “Insuficiencia” y “Elemental”, predominan en el campo de las matemáticas con 16.9% y 48,2% respectivamente; en el año 2018 - 2019, el promedio en dominio matemático es de 7,98 puntos sobre 10, es decir, se encuentra en un nivel de logro “Elemental”, considerando que el sector rural refleja un índice de deficiencia matemática de 0,17 en comparación al sector urbano. Además se ha demostrado en base a [2] que existe mejores resultados en instituciones particulares que en instituciones públicas (fiscales y fiscomisionales). Esto presentaría un inconveniente en las carreras del futuro debido al bajo potencial de conocimiento que se presenta en los aspirantes que intentan ingresar a la Universidad [3].

La matemática es una materia principal y elemental dentro de la formación de escuelas y colegios, destacando su importancia en las carreras de formación universitaria ya que su aplicación se ve vinculada en su ámbito laboral. Actualmente constituyen uno de los conocimientos imprescindibles en las sociedades modernas con un desarrollo tecnológico muy importante. Sin embargo, se observa en los diferentes niveles educativos del Ecuador, que las matemáticas se presentan como uno de los conocimientos poco comprendidos para muchos estudiantes, esto podría orillar al estudiante a abandonar sus estudios [4].

La deserción estudiantil afecta al sistema educativo superior en todo el país, de acuerdo a la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) se alcanza el 40 % de estudiantes que abandonan la universidad, donde el mayor índice de deserción ocurre en los primeros semestres [5]. Según [5-10] hacen referencia a los factores más comunes para que los estudiantes deserten a: factores personales, académicos y socioeconómicos. Mientras que [11] determina que existen factores pedagógicos (metodología de enseñanza-aprendizaje) que se involucran en el abandono de los estudiantes.

Por otra parte, en Loja, las universidades: Universidad Técnica Particular de Loja (UTPL), la Universidad Nacional de Loja (UNL) y la Universidad Internacional del

Ecuador (UIDE) presentan una semejanza entre las carreras más demandadas a “Medicina Humana”, “Administración de Empresas” y “Derecho”<sup>1</sup>, dejando a un lado las carreras de los campos de conocimiento vinculados con las carreras STEAM como: “Ingeniería, Industria y Construcción”, “Tecnologías de la Información y Comunicación”, “Ciencias Naturales, Matemáticas y Estadística”<sup>2</sup>. Añadiendo que dentro de lo antes mencionado existe desigualdad de género, debido a que el número de mujeres dentro de las ingenierías a nivel del país<sup>3</sup> es únicamente del 10%. Además, según<sup>4</sup> los hombres tienen un rendimiento de 20 puntos superior al de las mujeres, tomando en cuenta que un año escolar es representado por 30 puntos, en consecuencia las mujeres tendrían un año de escolaridad menos a de los hombres en Ecuador.

De acuerdo a entrevistas aplicadas a docentes y estudiantes de las Carreras de Ingeniería en Sistemas y Computación (Ver Anexo I), docente y estudiantes universitarios coinciden que la falta de bases matemáticas adquiridas en los colegios no se ajusta a los requerimientos de las asignaturas que la carrera necesita. De igual forma, se aplicó entrevistas a 29 alumnos de 3 de BGU paralelo “D” del Colegio “Bernardo Valdivieso”, en donde 10 estudiantes informaron que presentan inconvenientes o les desagrada las matemáticas (ver Anexo I), esto se ve reflejado en los resultados de “PISA\_D” y examen “Ser bachiller”. Por otra parte, a través de la misma se logró identificar que 13 de 29 estudiantes muestran interés por la carrera de “Medicina Humana”, esto perteneciente a 11 mujeres y 2 hombres; 5 alumnos entre ellos 3 mujeres y 2 hombres se inclinan por “Derecho”. Mientras tanto, 5 estudiantes todos hombres buscan seguir una ingeniería, los demás, hace referencia a otras carreras tales como “Filosofía”, “Veterinaria”, entre otras. Al comparar estas evidencias se sigue reflejando el mismo comportamiento por parte de los estudiantes de dejar a un lado las carreras tecnológicas o científicas, tomando en cuenta que el número de mujeres dentro de las mismas es bajo. Por su parte, estudiantes universitarios tanto hombres como mujeres indican que la carrera que actualmente cursan no fue su primera alternativa dentro de las postulaciones que brinda la SENESCYT. En consecuencia, existe el bajo rendimiento académico, la desmotivación, incluso puede provocar el abandono de la carrera. Lo que da origen a la siguiente pregunta de investigación:

---

<sup>1</sup>“Medicina y derecho son las carreras con más demanda en Loja: Noticias Loja: La Hora Noticias de Ecuador, sus provincias y el mundo.” [Online]. Available: <https://lahora.com.ec/loja/noticia/1102140041/medicina-y-derecho-son-las-carreras-con-mas-demanda-en-loja->.

<sup>2</sup>“El Telégrafo - Noticias del Ecuador y del mundo - Apenas el 10% de mujeres en el país se inclina por las ingenierías.” [Online]. Available: <https://www.eltelegrafo.com.ec/noticias/sociedad/6/apenas-el-10-de-mujeres-en-el-pais-se-inclina-por-las-ingenierias>.

---

## 1.2. Problema de Investigación

¿El uso de técnicas de minería de datos educativa permitirá identificar los patrones que existen en los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas en la carrera de Ingeniería en Sistemas/Computación?

---

## 2. Justificación

El presente Trabajo de Titulación (TT) se relaciona con dos objetivos de Desarrollo Sostenible (ODS), “Educación de Calidad” y “Equidad de Género”; en las áreas de investigación priorizadas por el SENESCYT, el proyecto se enmarca en la línea de “Tecnologías de la Información y Comunicación (TIC)” y sub-líneas “Ciudades inteligentes e inclusivas”, “Economía, tecnología y sociedad”, y “TIC para la educación e inclusión social”. De la misma manera dentro de la línea de “Territorio y Sociedad”, vinculado a: “Estudios sobre Género y violencia de género” y “TIC aplicada al territorio y la inclusión”. Por otra parte, dentro de: “Libro Blanco de la Sociedad de la Información y del Conocimiento (2018)” y “Libro Blanco de las Líneas de Investigación, Desarrollo e Innovación y Transferencia del Conocimiento en TIC (2019)” el proyecto se ve vinculado en las sub-líneas de “Inclusión y Habilidades Digitales” y “TIC y Educación” respectivamente. Asimismo, el proyecto tendrá un impacto en las líneas de Investigación de Educación, Cultura y Comunicación; Energía, Industrias y Recursos Naturales no Renovables que busca promover y organizar las actividades de investigación científica dentro de la Universidad Nacional de Loja, y, dentro de las líneas de investigación de la carrera de Ingeniería en Sistemas, el proyecto se relaciona directamente con Sistemas Inteligentes e Informática Educativa Inteligente.

En Ecuador existe la organización denominada “Coalición STEM”<sup>1</sup>, conformada por el Ministerio de Educación, la Red Ecuatoriana de Mujeres Científicas, entre otras instituciones que buscan contribuir al mejoramiento de la formación de docentes en matemáticas, ciencias y tecnología; así como también erradicar la ansiedad matemática de los niños y niñas. Según [12], la ansiedad en el aprendizaje de las matemáticas afecta de manera uniforme al desarrollo de la cognición matemática. Algunos individuos que tienen altos niveles de ansiedad matemática son más evasivos, considerando que el rendimiento matemático no sólo varía en función de dicha ansiedad, sino también en cómo las personas se acercan a situaciones relacionadas con las matemáticas. En referencia a lo antes mencionado, el TT servirá de apoyo para cumplir el objetivo de la organización, esto implica mejorar la manera en como los estudiantes reciben el conocimiento por parte de los docentes y sobretodo lograr incrementar la cantidad de estudiantes que consideran una carrera en un campo relacionado con STEM.

El desarrollo del presente TT servirá como recurso de apoyo científico, tecnológico y académico a investigaciones que se vinculen a la misma línea de investigación o se

---

<sup>1</sup>Coalición STEM.” Available: <http://www.stem.ec/>.

extrapole hacia un tema de estudio en particular. De la misma manera, permitirá a las autoridades correspondientes elaborar y ejecutar estrategias con el propósito de mejorar la toma de decisiones que permitan disminuir los altos niveles de deserción estudiantil, factores que afectan el rendimiento académico de los estudiantes y reducir los gastos económicos que se generan al estado. Además, la aplicación práctica del TT se da por la posibilidad que ofrece investigadores, profesores y estudiantes un vínculo con la educación STEM como fuente que permite formar profesionales para el futuro.

---

## 3. Objetivos

### 3.1. General

Aplicar técnicas de Minería de Datos Educativa para identificar patrones en los registros académicos de los estudiantes con asignaturas relacionadas a las matemáticas en la Carrera de Ingeniería en Sistemas/Computación de la Universidad Nacional de Loja.

### 3.2. Específicos

- Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería.
- Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.
- Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.

## 4. Alcance

Para identificar los patrones en los registros académicos de los estudiantes de la Carrera de Ingeniería en Sistemas/Computación de la UNL con asignaturas relacionadas a las matemáticas, el TT se desarrollará en base a las actividades detalladas a continuación con el propósito de cumplir con los objetivos planteados.

Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería.

- Establecer métricas y lineamientos para la selección de trabajos relacionados con el tema de estudio.
- Obtener y analizar los trabajos relacionados en base a métricas y lineamientos antes mencionados.
- Elaborar Documentación.

Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de ingeniería por medio de Minería de Datos Educativa.

- Seleccionar la herramienta de minería de datos a utilizar para el desarrollo del trabajo.
- Determinar el algoritmo más óptimo para la minería de datos.
- Analizar e interpretar los resultados obtenidos.

Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.

- Definir los objetivos del plan de acción.
- Construir el plan de acción de acuerdo a los objetivos planteados.
- Elaborar el plan de acción.

# 5. Marco Teórico

## 5.1. Deserción Estudiantil

El vocablo deserción proviene del latín “desertio”, que significa abandono. Aplicado al ámbito educativo, la deserción es considerada como el abandono de la formación académica en el sistema de educación formal, independientemente de las condiciones y modalidades presenciales, debido a causas internas y externas que intervienen en el mismo [13]. La deserción estudiantil es uno de los problemas que aborda la mayoría de las instituciones de educación superior de toda Latinoamérica. A través de distintas investigaciones, se da cuenta que la interrupción de los estudios universitarios, antes de llegar a su finalización, es un fenómeno generalizado que ha afectado a varias entidades educativas [14].

### 5.1.1. Deserción Universitaria en Ecuador

La UNESCO es uno de los pocos organismos que ha realizado estudios globales acerca del problema de deserción universitaria y las consecuencias que lleva consigo en la economía de los países. Según [5] indica que en la universidad existen estudiantes que terminan las asignaturas pero no es necesario para obtener el título, debido a las dificultades que presentan en la culminación de su proyecto de grado, estos demuestran los vacíos existentes en el desarrollo de habilidades en lectura y escritura no solo se evidencian en la educación básica y media sino que continúan en la universidad.

### 5.1.2. Factores que causan la deserción Universitaria

En estudios realizados en distintas ciudades de Ecuador, se ha logrado determinar que existe un número elevado de deserción estudiantil debido a la alta exigencia académica de parte de los docentes, los mismos que no cuentan con una preparación pedagógica adecuada que permita motivar, orientar y guiar a los estudiantes para que continúen sus estudios [9].

Según [4] hace referencia a ciertos factores relacionados directamente con el miedo, temor que generan las matemáticas en los diferentes niveles de educación. 1) La influencia del docente que imparte la materia. Esto implica tener en cuenta la impotación de la motivación dentro del sistema educativo como un aporte de los docentes hacia los



estudiantes con el propósito de evitar ese temor a las matemáticas. 2) El rol de la familia durante el proceso de aprendizaje del estudiante y, 3) El aspecto socioeconómico del estudiante y la repercusión que éste ocasiona en el rendimiento académico del alumno.

De manera general y como aporte, en base a los resultados generados por estudios de la Unesco y en relación a [5, 11, 14], se determinan algunos factores comunes dentro de la deserción estudiantil universitaria a nivel del País, tales como:

#### **Factores Socioeconómicos**

- Nivel socioeconómico del estudiante
- Situación laboral del estudiante
- Personas a cargo del estudiante
- Nivel educativo de los padres

#### **Factores de Riesgo Académicos**

- Orientación profesional
- Tipo de colegio en nivel medio
- Rendimiento académico
- Métodos de estudio
- Calificaciones en el examen de admisión
- Insatisfacción con el programa universitario
- Carga académica

#### **Factores de Riesgo Individuales**

- Edad, género y estado civil
- Entorno familiar
- Integración social
- Incompatibilidad horaria con actividades extra académicas

#### **Factores de Riesgo Institucionales**

- Normativas académicas
  - Becas y financiamientos
-

- Recursos universitarios
- Relaciones con profesores y estudiantes
- Grado de compromiso con la universidad

Cada uno de los factores descritos anteriormente presentan una similitud ante el problema principal que se presenta en las diferentes instituciones de Educación superior considerando que todo se vincula con factores personales, académicos, institucionales y económicos. Existen diferentes motivos que pueden orillar a un estudiante a abandonar sus estudios, es por ello que se debe tomar en cuenta cada acción que sirva como indicador permitiendo de esta manera evitar el abandono de los alumnos a los largo de sus estudios.

## 5.2. Minería de Datos

La Minería de Datos o Data Mining es “un conjunto de técnicas y herramientas aplicadas al proceso de extraer y presentar conocimiento implícito, previamente desconocido, útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir, de forma automatizada, tendencias o comportamientos” [13].

Estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos; además se pretende que la información obtenida contenga capacidad predictiva, facilitando de esta manera el análisis de los datos de manera eficiente. Bajo la denominación de “minería de datos” se han agrupado recientemente diversas técnicas estadísticas y del aprendizaje automático (Inteligencia Artificial) enfocadas, principalmente, a la visualización, análisis, y modelización de información de bases de datos masivas [15].

### 5.2.1. Minería de Datos Aplicada al campo Educativo

La Minería de Datos Educativa (EDM) es una disciplina emergente, que se interesa por el desarrollo de métodos para la exploración de las características singulares de los datos que provienen de entornos educativos, y de esta manera utilizar esos métodos para comprender mejor el desempeño de los estudiantes, y las condiciones en las cuales ellos aprenden más. Es el proceso de transformar los datos en bruto, recopilados por los sistemas de enseñanza, en información útil que pueda utilizarse para tomar decisiones y responder preguntas de investigación. La aplicación de la MD en el ámbito de la enseñanza, tiene como objetivo obtener una mejor comprensión del proceso de aprendizaje de los estudiantes y de su participación global en el proceso, orientado a la mejora de la calidad y la eficiencia del sistema educativo. A partir de toda la información

---

disponible, las diferentes técnicas de MD pueden ser aplicadas a fin de descubrir conocimiento útil que ayude a mejorar el proceso educativo, siendo este conocimiento muy diverso el mismo que permitirá determinar los factores que acompañan al estudiante en el proceso educativo [16].

La minería de datos educativa ofrece numerosas ventajas comparándola con los paradigmas más tradicionales de investigación relativa a la educación, como experimentos de laboratorio, estudios sociológicos o investigación de diseño [17]. En el sector educativo, las técnicas de minería de datos son usadas para la comprensión del comportamiento de los estudiantes, la Minería de Datos Educativa emerge como un paradigma orientado para la generalización de modelos, tareas, métodos y algoritmos para la exploración de datos que provienen de un contexto educativo, de la misma manera tiene como función encontrar, determinar, analizar patrones que caractericen los comportamientos en base a sus logros, evaluaciones y el dominio de contenido de conocimiento que tienen los alumnos en los diversos mecanismos de aprendizaje-enseñanza que hoy en día son otorgados en las diversas instituciones públicas y privadas con el objetivo de generar modelos educativos en los cuales puedan fomentar nuevas estrategias, técnicas o herramientas que puedan analizar e incrementar el nivel participativo, educativo e interactivo de los estudiantes sobre los sistemas de aprendizaje-enseñanza que las diferentes instituciones presentan [18].

La Minería de Datos Educativa - EDM es un área multidisciplinaria en que intervienen distintos paradigmas de computación como son el desarrollo o construcción de algoritmos tales como: predicción, programación lógica, algoritmos estadísticos, entre otros, con el propósito de generar tareas principales como; la clasificación, agrupamiento (Clustering), estimación, modelado de dependencias, visualización y descubrimiento de reglas con el fin de construir un modelo que se ajuste a un conjunto de datos sobre un contexto educativo, determinando de esta manera un conocimiento certero del sistema y predecir comportamientos futuros del mismo [13, 18].

### 5.3. Trabajos Relacionados

Para examinar el estado de la cuestión sobre la identificación de patrones con asignaturas relacionadas a las matemáticas en información de estudiantes de Ingeniería, se busca obtener una base sólida de literatura científica, la misma que se ha identificado por medio de una Revisión Sistemática de Literatura, la cual permitirá identificar los trabajos relacionados que servirán de apoyo para la elaboración del TT, considerando que se ha realizado una referencia textual de los mismos. La Tabla 5.1 refleja los 19 documentos o trabajos relacionados encontrados, de los cuales TR-01 al TR-14 fueron obtenidos en base a la RSL<sup>1</sup>, y desde TR-15 hasta TR-19 son documentos extraídos del repositorio Dspace de la Universidad Nacional de Loja.

---

<sup>1</sup>El proceso para la determinación de los trabajos relacionados se ejecutó en base a la revisión sistemática de literatura la misma que se encuentra a detalle en el Anexo II.

---

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-01	Minería de datos aplicada a estrategias para minimizar la deserción universitaria en carreras de Informática de la UNNOBA [13].	Es una investigación desarrollada en el campo de la Ciencia Informática aplicada en Educación, que busca a través de minería de datos definir y desarrollar un modelo de intervención áulica virtual que permita diseñar, e implementar, estrategias tecnológicas tendientes a realizar un seguimiento de las trayectorias educativas de aquellos estudiantes con rezago académico o riesgos de deserción debido a que los índices de deserción y desgranamiento en estas carreras son muy elevados sobre todo en los dos primeros años. Para llevar a cabo la investigación se hace uso de algoritmos de clasificación, árboles de decisión, entre otros presentes en el software de la institución
TR-02	Estilos de aprendizaje y minería de datos: un estudio preliminar en el contexto universitario [19]	En base a una investigación, se presenta los resultados sobre los estilos de aprendizaje y rendimiento académico en estudiantes de ingeniería, para llevar a cabo la investigación se hizo uso de minería de datos como técnica para el análisis de datos. Lo que busca es evitar la deserción estudiantil y a su vez mejorar las prácticas pedagógicas de los docentes con el único objetivo de mejorar el rendimiento académico y ofrecer a los docentes estrategias pedagógicas para mejorar su práctica en el aula, en su mayoría los estudios se quedan en una fase descriptiva.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-03	Aplicación de técnicas de minería de datos para la identificación de patrones de deserción estudiantil como apoyo a las estrategias de SARA (sistema de acompañamiento para el rendimiento académico) [20].	Hace referencia a la deserción estudiantil en instituciones de educación superior, con la perspectiva de elaborar modelos de minería de datos los mismos que identificar patrones de deserción. Se realizó una RSL para tener una visión más precisa sobre los factores de deserción universitaria. Para ello se identificó el problema concretamente en el Programa de Ingeniería de Sistemas y Computación de la Universidad del Quindío. Además, se presentan causas de deserción comunes entre los estudiantes, así mismo, se detalla el Proyecto SARA (Sistema de Acompañamiento para el Rendimiento Académico), el mismo que busca ayudar y acompañar académicamente a los estudiantes en los primeros semestres o aquellos que presentan bajo rendimiento académico, brindando apoyo en los temas o áreas en los que el estudiante presente dificultad de aprendizaje. De la misma manera busca analizar las causas de deserción y con las mismas crear nuevas estrategias que eviten o ayuden a contra restar el problema. incorporar la aplicación de un Proceso de descubrimiento de conocimiento en bases de datos o KDD, para caracterizar y predecir posibles casos de deserción en el programa a través de modelos de minería de datos. El conocimiento de los patrones permitirá generar e incluir estrategias alineadas con la visión y los objetivos que persigue SARA.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-04	Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos [21].	Se presenta un análisis acerca de la deserción de los estudiantes a través de técnicas de minería de datos y con el mismo obtener un modelo con la capacidad de clasificar a estudiantes que desertaran tomando en cuenta datos socioeconómicos y académicos de los estudiantes. Este estudio se realizo bajo la metodología CRISP-DM, analizando tres modelos de clasificación: árboles de decisión, métodos bayesianos y redes neuronales, con el fin de evaluar su comportamiento, encontrándose que Random Forest es el algoritmo de mejor desempeño general, con un 88,9% de exactitud a diferencia del algoritmo Naive Bayes. Mediante los experimentos realizados se determinó que las variables académicas de ingreso no afectan significativamente la deserción de primer año.
TR-05	Causas de Deserción en el Ingreso a la Universidad: un Estudio de Caso [6].	Es una investigación que analiza las causas de deserción que se generan en el desarrollo del curso de nivelación previo al ingreso al primer año de las facultades del área técnica de la Universidad de Cuenca. Para identificar las principales causas de deserción se analizaron las fichas socioeconómicas de los estudiantes, se aplicaron pruebas de matemáticas al inicio y fin del curso, así como también una encuesta; también, se realizó una encuesta a una muestra de estudiantes que desertaron en la nivelación, y se analizaron tasas de deserción de estudiantes que cursan las carreras del área técnica. Los resultados reflejan algunas causas de deserción como: la carencia de bases en temas de matemáticas, no entrar a la carrera de su preferencia, alta exigencia académica en carreras técnicas, un sistema de evaluación complejo, falta de recursos económicos, entre otras razones para la deserción. Algunos de estos factores se relacionan con otros estudios realizados donde se pretende reducir el índice de deserción estudiantil universitaria.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-06	Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos [22].	Se analiza información académica con el objetivo de identificar factores que influyen sobre la deserción de los estudiantes de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina, mediante la aplicación de una técnica de minería de datos. La fuente de datos contiene información proporcionada al ingreso (personales y antecedentes educativos) y la que se genera durante el periodo de estudios. Se realiza la selección y depuración de datos, utilizando diferentes criterios de representación y aplicación de algoritmos de clasificación como árboles de decisión, redes bayesianas y reglas. Se identifica como variables influyentes en la deserción, asignaturas aprobadas, cantidad y resultado de asignaturas cursadas, procedencia y edad de ingreso del estudiante. Mediante este proceso fue posible identificar los atributos que caracterizan a los casos de deserción y su relación con el desempeño académico, especialmente en el primer año de la carrera.
TR-07	Predicción de deserción y éxito en estudiantes. Caso de estudio: ingeniería agroindustrial de la UNI Norte, Nicaragua, 2011-2015 [23].	Parte de una investigación en la que para poder buscar patrones de rendimiento académico, se hizo uso de estadística univariada y multivariada, la misma que permitiría de forma temprana predecir la deserción y el éxito, al final de la carrera, en 7 generaciones de estudiantes universitarios de ingeniería agroindustrial en la UNI sede Norte, Nicaragua. Por medio del análisis discriminante, el cual fue aplicado a las calificaciones del primer semestre de primer año de la carrera, se pudo clasificar los estudiantes que llegan a quinto año como límite ya sea de forma exitosa o no, con un alto nivel de acierto, 79%. Las matemáticas en primer lugar considerada como una de las importantes y luego las químicas fueron las asignaturas de mayor peso para predecir el éxito o fracaso final en la carrera dentro de la investigación.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-08	Estrategia pedagógica con Aprendizaje basado en Juegos-GBL, para fomentar la motivación en el Área de Matemáticas en Ingeniería [24].	El artículo propone una estrategia pedagógica, fundamentada en el aprendizaje basado en juegos-GBL, lo que permitirá incrementar la motivación de los estudiantes del área de matemáticas de primer año de ingeniería. Lo que se pretende mediante la estrategia es disminuir los índices de reprobación de los cursos de matemáticas. Para ello, se determinó como muestra de una población a un grupo de estudiantes de primer año de Cálculo Diferencial en Ingeniería, con un total de 19 estudiantes. Se propuso una metodología de investigación mixta, y mediante la misma se logró determinar o encontrar evidencias que las acciones formativas, mediadas por GBL, aportan en la motivación de los estudiantes y en su interés por los cursos de matemáticas. La planeación detallada, las actividades con un contexto claro y la robustez de las herramientas tecnológicas; son aspectos claves en el éxito de la experiencia dentro de las ingenierías.



**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-09	Rendimiento académico, deserción y competencias matemáticas en ingeniería: caso Escuela de Ingeniería de Sistemas, UNSA [25].	Indica el inconveniente que sufre los sistemas universitarios como lo es el caso de la Escuela de Ingeniería de Sistemas de la Universidad Nacional San Agustín, se enfrenta a un bajo rendimiento, deserción y rezago estudiantil, esto se ve reflejado en las asignaturas de matemáticas, lo que provoca un porcentaje considerable de estudiantes que no culminan sus estudios universitarios lo que provoca pérdidas económicas y personales. Para poder determinar con exactitud el problema reflejado en la institución educativa que tomó en cuenta las calificaciones y con base a las mismas, un conjunto de indicadores que tratan de describir el rendimiento académico de los estudiantes. La investigación explica la importancia de las matemáticas en la formación de un ingeniero en sistemas, de esta manera a lo largo de su profesión se podrá detectar el problema y a su vez analizar las ventajas determinadas a lo largo del proceso, conjuntamente con la formación del estudiante universitario.
TR-10	Brecha de género en matemáticas: el sesgo de las pruebas competitivas [26].	Hace referencia a la brecha de género en las matemáticas, existiendo un índice de diferencia radical entre hombre y mujeres en resultados de PISA 2012. Los bajos resultados en matemáticas de las mujeres afectan no solo su acceso al derecho a la educación, sino que también impactan en la menor participación de las mujeres en la matrícula de las carreras del área de Ciencia, Tecnología, Ingeniería y Matemáticas (STEM), tomando en consideración que estas carreras presentan más oportunidades laborales. Para poder analizar esta brecha de género se realiza un test de matemáticas que logrará identificar si realmente se mantiene esa brecha o puede verse afectado por otros factores. Sin embargo en los resultados se identifica que hay mujeres que logran romper con la brecha y logran grandes niveles de éxito superior al de los hombres.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-11	Estrategia para afrontar la deserción universitaria desde las tecnologías de la información y las comunicaciones [27].	Es un artículo que busca describir los principales componentes de una estrategia a través de las TIC para abordar el problema de deserción de estudiantes en este caso de pregrado. Lo que pretende es evitar la duplicación de esfuerzos y gastos de recursos al determinar si un estudiante está en riesgo de deserción. Todo esto se llevó a cabo mediante la heurística y el área de proyectiles, donde se determinó fases que permitan desarrollar el proyecto. Esto se ve como base para implementar nuevas estrategias que ayuden a evitar la deserción de los estudiantes.
TR-12	Modelo Predictivo Para La Identificación De Patrones De La Deserción Estudiantil En La Universidad Nacional Tecnológica De Lima Sur (Untels) [28].	Es un trabajo en el pretende determinar los patrones del entorno que impactan en la deserción de estudiantes de una universidad de Lima (Untels), para ello se elabora una base de datos con dato socioeconómico y académicos de un determinado rango de estudiantes (2007-2011). Para la investigación se elaboran seis modelos utilizando Decisión Tree de RapidMiner, con o sin validación cruzadas, con parámetros modificados. Cada modelo es comparado en base a su precisión. También se implementa J48, como algoritmo de la herramienta Weka. El modelo final detectó las siguientes características o patrones del entorno que impactan en la deserción de los estudiantes de la Untels: Número de matrículas en los cuatro semestres consecutivos a su ingreso, Promedio en su segunda matrícula, Edad de ingreso, Promedio en su cuarta matrícula, Año de ingreso, Número de personas dependientes, Semestre de ingreso y Número de cursos aprobados en su primera matrícula. Dando como principal patrón detectado para los desertores, es que el número de matrículas en los cuatro semestres consecutivos a su ingreso sea menor o igual que 3, el mismo con una precisión de 88%.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-13	Quality Improvements in Online Education System by Using Data Mining Techniques [29].	Aplica un método de enseñanza en línea, indica que el método tradicional es así de fácil de llevar debido a que permite al docente llevar un control de rendimiento del estudiante. Por su parte el aprendizaje en línea se dificulta monitorear el desempeño del estudiante, ante esto se realiza un sistema de predicción de deserción el mismo que seguirá el progreso del estudiante en el curso, además de funcionar como un sistema alerta para que el docente mejore el rendimiento del estudiante. Para poder llevar a cabo este sistema se hizo uso de algoritmos de agrupación como Random Forest, C4.5 presentes en el software Weka.
TR-14	Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout [30].	Presenta un enfoque computacional utilizando la minería de datos educativos y diferentes técnicas de aprendizaje supervisado como son: Árboles de Decisión, Vecino Cercano, Redes Neuronales, Máquinas Vectoriales de Apoyo, Bayes Ingenuas y Bosques Aleatorios, con el objetivo de evaluar el comportamiento de diferentes modelos de predicción con el fin de identificar el perfil de los estudiantes universitarios en riesgo de deserción. En los resultados aplicando los diferentes algoritmos se logró identificar que los estudiantes abandonaron sus respectivos programas con mayor frecuencia en el semestre 4 para los programas de Informática e Sistemas de Información, y en el semestre 6 en Ingeniería Informática, ya que estos fueron los semestres en los que los algoritmos tuvieron mejores resultados. Con esto se logra identificar que implementar herramientas para determinar los factores de deserción en los estudiantes es fundamental para poder tratar de mitigar dicho inconveniente.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-15	Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC [31].	Este es uno de los Trabajos de Titulación previo a la Obtención del título, el cual tiene por objetivo obtener la segmentación de clientes en la empresa tecnológica Master PC a través de la aplicación de técnicas de Minería de Datos, por ello se consideró variables como el comportamiento de compra, la misma que permitió identificar la lealtad de los clientes de la empresa tecnológica. Para el desarrollo del trabajo se aplicó la metodología CRISP-DM que sirvió de base durante todo el proceso de Minería de Datos. El análisis se realizó en base al modelo RFM (Recencia, Frecuencia, Valor Monetario), y sobre este modelo se aplicaron los algoritmos de agrupamiento: k-means, k-medoids, y Self-Organizing Maps (SOM). Para la validación del resultado de los algoritmos de agrupamiento y seleccionar el que proporcione grupos de mejor calidad, se ha aplicado la técnica de evaluación en cascada aplicando un algoritmo de clasificación. Como apoyo se utilizó el algoritmo Apriori para encontrar asociaciones entre productos, para cada grupo de clientes. Se utilizó el entorno RStudio para el proceso de Minería de Datos.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-16	Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje. [32].	El presente trabajo se encuentra enfocado en determinar las interacciones de los estudiantes del curso virtual de inglés de la Modalidad de Estudios a Distancia (MED) de la Universidad Nacional de Loja, para ello se realizó un análisis de la base de datos donde se encontraba la información correspondiente al periodo académico 2013 - 2014, que consta de datos personales, institucionales, socioeconómicos y de las interacciones (tareas, recursos y exámenes) de los estudiantes. De igual forma se consideró realizar un estudio de las diferentes técnicas de minería de datos acordes al proyecto donde se seleccionó la técnica de clasificación para generar los modelos a través de los algoritmos Chaid, Prism, Knn, Decision Tree, ID3, J48, Jrip y Ridor. Para el desarrollo del proyecto se consideró la metodología CRISP-D en base a un análisis de las metodologías de minería de datos comparando cada una de ellas con el fin de seleccionar la que más se acople al tema de estudio; esta contiene múltiples fases indicando cada una de sus actividades que se deben cumplir para obtener el modelo, convirtiéndose de esta forma en una guía práctica para cumplir con los objetivos establecidos. Además se desarrolló un análisis comparativo tomando en cuenta características de las herramientas de minería de datos donde se eligió RAPIDMINER para realizar los procesos mediante los algoritmos conjuntamente con los datos de los estudiantes los mismos que se dividieron en dos conjuntos, para entrenamiento y validación, obteniendo como resultado que el mejor algoritmo fue el Decision Tree, ya que clasificó las instancias correctamente así mismo presentó un árbol con los diferentes atributos dando las mejores reglas de las interacciones de los estudiantes de tal forma se pudo generar el modelo que permitió establecer que en gran mayoría los estudiantes tienen un nivel de interacción medio en el curso virtual de inglés, donde los factores que más influyeron fueron las interacciones en las tareas, exámenes, recursos, situación laboral y estado civil del estudiante.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-17	Identificación de factores en la reprobación y deserción mediante técnicas de minería de datos en el Área de la Energía de la Universidad Nacional de Loja. [33].	El objetivo principal del Trabajo es la identificación de factores que inciden en la deserción y reprobación universitaria. Para realizar todo este proceso, se aplicaron de técnicas de minería de datos. En el trabajo realizado se llegó a la conclusión de que el factor que incide mayormente en la deserción de los estudiantes es el académico, seguido por el institucional dejando por último al factor individual, mientras que en el ambiente de reprobación los estudiantes que reprueban presentan la siguiente condición: Si el estudiante tiene un promedio de notas regular, un promedio de asistencia bajo y pertenecen a la carrera de Ingeniería en Sistemas entonces si reprueba. Los mejores resultados se obtuvieron mediante los algoritmos basados en reglas como Ridor, Part y JRip. Con el desarrollo del presente trabajo se ha aprendido a seguir la metodología En cuanto a los métodos aplicados para lograr el desarrollo eficiente del presente trabajo se ha utilizado la metodología CRISP-DM, la cual da la posibilidad de llevar un trabajo ordenado e iterativo. Los datos utilizados para llevar a cabo la identificación de los factores, se obtuvo de las bases de datos del Sistema de Gestión Académica (S.G.A.) a través de su Web Services, además se recopiló datos del Área de Bienestar Universitario, todo esto para luego integrar en una sola fuente de datos, para su posterior procesamiento. Finalmente luego de identificar los factores de deserción y reprobación, se generó un modelo predictivo de deserción con el propósito de validar los resultados obtenidos, el cual se evaluó con datos de estudiantes que cursan actualmente las carreras del Área de Energía de las Industrias y los Recursos Naturales No Renovables de la Universidad Nacional de Loja.

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-18	Determinación de perfiles profesionales mediante técnicas de minería de datos. [34].	<p>En presente trabajo de titulación desarrollado se evidenció que el perfil profesional que más predomina en los últimos egresados del año 2014 es desarrollador de software cuyos conocimientos obtenidos en las aulas universitarias serán puestos en práctica en el desarrollo de su vida profesional, ésta información es muy útil para los futuros cambios que se realicen a nivel de la malla curricular y perfil de carrera. Para el desarrollo del trabajo se implementó la metodología CRISP-DM, la misma que ayudó a organizar mediante fases, sub fases y tareas que apoyaron a la documentación del proyecto a más de ser una guía para el desarrollo durante todo el proceso. Los algoritmos ID3 y CHAID que pertenecen a las técnicas de clasificación basadas en árboles de decisión y los algoritmos JRip, PART, Ridor, Decisión Table, DTNB y NNge, pertenecientes al grupo de técnicas de reglas de inducción. Ya en el desarrollo y generación de modelos, los mejores algoritmos fueron CHAID y JRip los cuáles se hicieron con el 72% de los datos y con el 28% restante se hizo la evaluación de los mismos para verificar su validez, donde CHAID resultó el más óptimo al clasificar el 96.55% de las instancias; mientras que JRip clasificó el 91,38%. Posterior a ello se realizó la aplicación de éstos algoritmos en un contexto real para validar y realizar sí la elección final, en donde JRip tuvo el mejor rendimiento en la predicción con el 100%; mientras que CHAID realizó la predicción del 76%, llegando a la conclusión que JRip es el modelo que se debe aplicar para la obtención de los perfiles profesionales.</p>

**Tabla 5.1:** Trabajos Relacionados

N°	Título del proyecto	Descripción
TR-19	Aplicación de minería de datos para determinar las causas de la Violencia de género contra las Mujeres en Ecuador [35].	Este trabajo se enfocó en los casos registrados en el Ecuador por el INEC mediante la Encuesta Nacional de Relaciones Familiares y Violencia de Género contra las mujeres, que reposa en su sitio web. A esta base de datos le fueron aplicadas técnicas de minería de datos con la finalidad de determinar las causas principales por las que una mujer sufre violencia, para ello fueron considerados el ámbito de violencia en la pareja, violencia durante el noviazgo, violencia en la niñez frente a la violencia de pareja o noviazgo, y violencia en el ámbito social, laboral y educativo. Tomando en cuenta cada uno de estos ámbitos, se fragmentó la base de datos obteniendo treinta subconjuntos de datos orientados a cada tipo de violencia, ámbito en el que ocurrió y considerando el estado civil de la mujer. Se definieron cuatro tipos de violencia: psicológica, física, económica o patrimonial y sexual, de la misma manera fueron seleccionados seis algoritmos de clasificación para el minado de datos: J48, CHAID exhaustivo, redes neuronales, IBk, tabla de decisión y Random forest; las herramientas de software empleadas fueron RStudio, SPSS Statistics y Weka. En algunos casos fue necesaria la aplicación del algoritmo SMOTE a los conjuntos de datos, con la finalidad de optimizar los resultados. Para el desarrollo del proyecto se hizo uso de cinco fases de la metodología KDD.



## 6. Metodología

El presente TT se desarrollará en la Universidad Nacional de Loja, en la Facultad de Energía, los Recursos, las Industrias y los Recursos Naturales No Renovables, dentro de la Carrera de Ingeniería en Sistemas/Computación. Donde se utilizarán métodos y técnicas pertinentes para la obtención de información relevante, y de esta manera cumplir con los objetivos planteados.

Examinar el estado de la cuestión sobre el tema de estudio en Carreras de Ingeniería.

- **Revisión Bibliográfica.**- permitirá conocer, seleccionar y recopilar información acerca de la deserción universitaria con la aplicación de técnicas de Minería de Datos Educativa en investigaciones similares, además de servir de apoyo para la sustentación teórica de la Investigación, basándose en consultas de fuentes bibliográficas confiables.

Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.

- **Estudio de casos.**- Será utilizado para realizar una exploración de los componentes que intervienen en el caso de estudio del presente TT, permitiendo de igual forma analizar situaciones y casos similares que serán de apoyo en el desarrollo del mismo.
- **Método Inductivo.**- ayudará a comprender el modelo de negocio para poder determinar los diferentes factores que influyen en la deserción estudiantil en las carreras de Ingeniería en Sistemas y Computación.

Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.

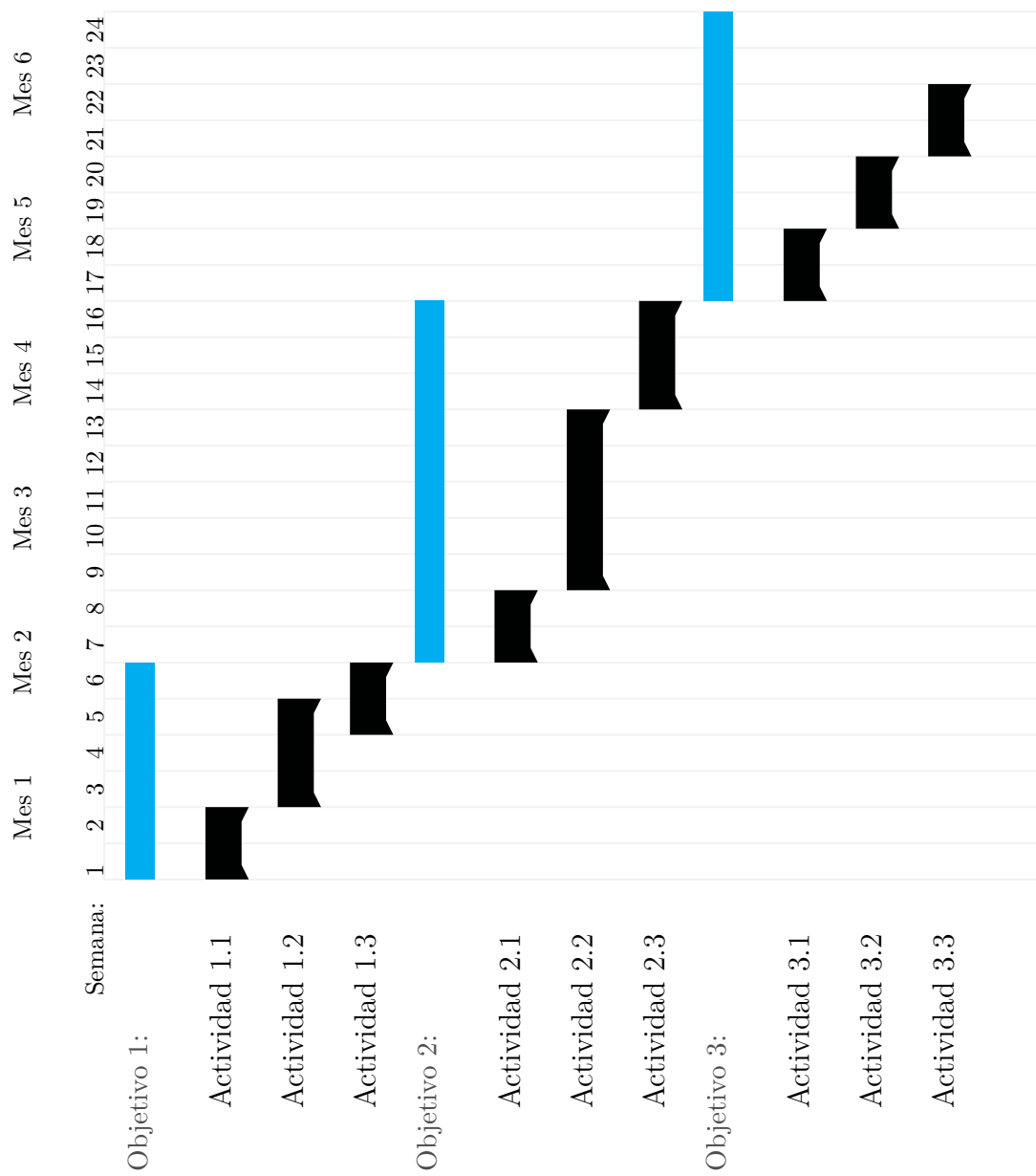
- **Observación Activa.**- Permitirá obtener una mayor comprensión sobre el problema de estudio dentro de las Carreras, buscando de esta manera determinar una posible alternativa acorde a las necesidades encontradas (resultados).
- **Entrevista.**- Ayudará a obtener información directa, relevante y necesaria que servirá de apoyo principalmente para la elaboración del plan de acción con los requerimientos específicos de acuerdo a los resultados.

## 7. Cronograma

Para la elaboración del cronograma se ha elaborado Tabla 7.1 donde se identifica las diferentes etiquetas a utilizarse en relación a los objetivos al ser plasmados en el cronograma de actividades.

**Tabla 7.1:** Etiquetas

<b>Etiqueta</b>	<b>Descripción</b>
Objetivo 1	Examinar el estado de la cuestión sobre la deserción de estudiantes universitarios en carreras de ingenierías.
Actividad 1.1	Establecer métricas y lineamientos para la selección de trabajos relacionados con el tema de estudio. Obtener y analizar de trabajos relacionados en base a métricas y lineamientos seleccionados. Elaborar Documentación
Actividad 1.2	
Actividad 1.3	
Objetivo 2	Identificar cuáles son los patrones predominantes en los registros académicos de los estudiantes de Ingeniería por medio de Minería de Datos Educativa.
Actividad 2.1	Seleccionar la herramienta de minería de datos a utilizar para el desarrollo del trabajo. Determinar el algoritmo para la minería de datos educativa. Análizar e interpretar los resultados obtenidos.
Actividad 2.2	
Actividad 2.3	
Objetivo 3	Proponer un plan de acción sobre los patrones predominantes identificados en los estudiantes de la Carrera de Ingeniería en Sistemas/Computación.
Actividad 3.1	Definir los objetivos del plan de acción. Construir el plan de acción de acuerdo a los objetivos planteados. Elaborar el plan de acción.
Actividad 3.2	
Actividad 3.3	



## 8. Presupuesto

Para la realización y culminación exitosa del presente TT se requiere realizar una inversión considerable, por lo que es muy importante realizar una estimación de costos, los mismos que se encuentran relacionados con cada una de las actividades y tareas que se necesitan para desarrollar el presente TT. Para que la estimación sea lo más acertada posible, se tomarán en cuenta varios aspectos, los cuales se detallan a continuación:

- **Talento Humano**

La Tabla 8.1 presenta una aproximación considerable sobre el costo por horas invertidas tanto por parte del investigador, como del Director y el Docente de la asignatura que se involucrarán en el presente TT.

**Tabla 8.1:** Presupuesto de Talento Humano

<b>Talento Humano</b>			
Equipo de Trabajo	Tiempo de trabajo (horas)	Precio(\$)/Hora	Valor Total(\$)
Investigador	400	3.00	1,200.00
Director del TT	50	10.00	500.00
Docente Materia de TT	120	10.00	1,200.00
<b>Total(\$)</b>			<b>2,900.00</b>

---

- **Recursos de Software y Hardware**

En esta sección se presenta la Tabla 8.2 la misma que contiene los costes aproximados de los recursos hardware, software y materiales de oficina que se utilizarán en el desarrollo del presente TT.

**Tabla 8.2:** Presupuesto de Software y Hardware

<b>Recursos de Software y Hardware</b>		
Rubro	Cantidad	Valor Total(\$)
Computadora	1	1,000.00
Pendrive	1	10.00
Impresora	1	350.00
Resma de Papel	1	3.50
Cartuchos de tinta	1	27.00
<b>Total</b>		<b>1,390.50</b>

---

- **Servicios**

En la Tabla 8.3 se muestran los servicios necesarios que se requerirán para el desarrollo del presente TT. En la misma se presenta el costo asociado a cada servicio tomando en cuenta el tiempo y la ocurrencia con la cual se requiere de estos servicios.

**Tabla 8.3:** Servicios

<b>Servicio</b>	
Rubro	Valor Total(\$)
Internet	120.00
Transporte	250.00
Copias	40.00
<b>Total</b>	410.00

- **Presupuesto General**

En la Tabla 8.4 se presenta la inversión aproximada que se realizará para el desarrollo y culminación exitosa del TT, esta tabla recopila cada uno de los costes involucrados de las diferentes secciones mencionadas en los apartados anteriores añadiendo un valor que cubrirá los posibles imprevistos que se pudiesen generar durante el desarrollo del TT.

**Tabla 8.4:** Presupuesto General

<b>Presupuesto General</b>	
<b>Rubro</b>	<b>Valor</b>
Talento Humano	2,900.00
Recursos de Software y Hardware	1,390.50
Servicios	410.00
Subtotal	4,700.50
Otros	470.05
<b>TOTAL(\$)</b>	<b>5,170.55</b>

Cabe recalcar que dentro del valor Total, lo que corresponde al presupuesto para el Director del Trabajo de Titulación y el docente de la misma, será patrocinado por la Universidad Nacional de Loja, siendo este un valor de \$1,700.00. Dicho esto, la diferencia del presupuesto estimado para la elaboración del TT con un valor de \$4,000.55, será financiado por el investigador.

# Bibliografía

- [1] INEVAL, “CIE\_ResumenEjecutivoPISA18\_20181123,” *Oecd*, p. 24, 2018.
- [2] G. M. López, A. C. Espinosa, V. M. López, and S. P. T. Cuenca, “Ser Bachiller 2018-2019,” p. 19, 2019.
- [3] L. Antonio, J. Luis, and J. Iván, “La educación STEM y su potencial en la formación integral del talento universitario para la Industria 4.0 en Ecuador.,” pp. 1–14, 2018.
- [4] J. C. Suárez, *Factores que generan miedo, apatía o desinterés frente al estudio de las Matemáticas*. PhD thesis, 2014.
- [5] G. J. Zambrano, K. G. Rodríguez, and L. H. Guevara, “Análisis de la Deserción Estudiantil en las Universidades del Ecuador y América Latina.,” pp. 1–28, 2018.
- [6] F. Bravo, L. Illescas, S. Larriva, and M. Peña, “Causas de Deserción en el Ingreso a la Universidad,” *Revista de la Facultad de Ciencias Químicas Universidad de Cuenca*, vol. 18, pp. 48–59, 2017.
- [7] E. V. García, R. C. Rodríguez, S. C. Tumbaco, Y. P. Santana, R. L. Fernández, and L. B. Benítez, “Factores que influyen en la deserción de los estudiantes en la Universidad de Guayaquil,” *Revista de la Facultad de Ciencias Médicas de la Universidad de Guayaquil*, vol. 0, no. 0, pp. 21–26, 2017.
- [8] L. d. C. Guevara, *Estudio de las causas que generan la deserción académica de los estudiantes de la escuela de Psicología*. PhD thesis, 2017.
- [9] N. S. Ortiz, X. F. Trujillo, and S. E. Tello, “Análisis de la deserción estudiantil en la carrera de ingeniería en telecomunicaciones para proponer nueva metodología de enseñanza,” *CISCI 2018 - Decima Septima Conferencia Iberoamericana en Sistemas, Cibernética e Informatica, Decimo Quinto Simposium Iberoamericano en Educacion, Cibernética e Informatica, SIECI 2018 - Memorias*, vol. 2, pp. 177–182, 2018.
- [10] D. A. Becerra, *Estudio del Rendimiento Académico Aplicando Técnicas de Minería de Datos*. PhD thesis, 2015.
- [11] A. Peña Fernández, “Los factores pedagógicos influyen en la deserción universitaria,” *INNOVA Research Journal*, vol. 4, no. 3, pp. 108–115, 2019.



- 
- [12] Z. Wang, S. L. Lukowski, S. A. Hart, I. M. Lyons, L. A. Thompson, Y. Kovas, M. M. Mazzocco, R. Plomin, and S. A. Petrill, “Is Math Anxiety Always Bad for Math Learning? The Role of Math Motivation,” *Psychological Science*, vol. 26, no. 12, pp. 1863–1876, 2015.
- [13] C. Russo and A. D. Giusti, “Deserción Universitaria En Carreras De Informática De La Unnoba,” *Universidad Nacional de la Plata*, 2019.
- [14] C. R. Passailaigue, O. Amechazurra, and J. Galarza, “La deserción y la repitencia en las instituciones de Educación Superior: algunas experiencias investigativas en el Ecuador.,” *Universidad y Sociedad*, vol. 6, no. 1, pp. 102–107, 2014.
- [15] M. B. Beltrán, “Minería de Datos,” p. 67.
- [16] K. Eckert and R. Suénaga, “Aplicación de técnicas de minería de datos al análisis de situación y comportamiento académico de alumnos de la UGD,” ... *de Investigadores en Ciencias de la ...*, pp. 92–96, 2013.
- [17] Á. Jiménez and H. Álvarez, “Minería de Datos en Educación,”
- [18] A. Ballesteros, D. Sánchez-Guzmán, and R. García, “Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo.,” *Latin-American Journal of Physics Education*, vol. 7, no. 4, pp. 662–668, 2013.
- [19] C. Vergaraz, A. Oviedo, and C. Carmona, “Estilos de aprendizaje y minería de datos : un estudio preliminar en el contexto universitario,” 2018.
- [20] L. C. Calvache Fernández, V. Alvarez Vallejo, J. I. Triviño Arbeláez, C. Quiceno Restrepo, and R. Pulgarin Giraldo, “Aplicación de técnicas de minería de datos para la identificación de patrones de deserción estudiantil como apoyo a las estrategias de SARA (Sistema de acompañamiento para el rendimiento académico),” *Octava conferencia Latinoamericana sobre el abandono en la Educación superior*, pp. 1177–1185, 2018.
- [21] C. Z. Torres, C. A. Ramos, and J. L. Moraga, “Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año , mediante minería de datos.,” *Ciencia Amazónica: (Iquitos)*, ISSN 2221-5948, ISSN-e 2222-7431, Vol. 6, N<sup>o</sup>. 1, 2016, págs. 73-84, vol. 6, no. 1, pp. 73–84, 2016.
- [22] K. B. Eckert and R. Suénaga, “Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos,” *Formacion Universitaria*, vol. 8, no. 5, pp. 3–12, 2015.
-

- 
- [23] L. M. Dicovskiy and M. E. P. Pacheco, “Predicción de deserción y éxito en estudiantes. Caso de estudio: Ingeniería Agroindustrial de la UNI Norte, Nicaragua, 2011-2015,” vol. 31, no. 01, pp. 16–27, 2018.
- [24] S. Zabala-Vargas, L. García-Mora, and B. Benito-Crosetti, *Estrategia pedagógica con aprendizaje basado en juegos-GBL, para fomentar la motivación en el área de matemáticas en ingeniería*. No. August, 2019.
- [25] N. Bedregal and D. Tupacyupanqui, “Rendimiento académico, deserción y competencias matemáticas en ingeniería: caso Escuela de Ingeniería de Sistemas, UNSA,” no. September, 2018.
- [26] Ó. Arias R., A. Mizala S., and F. Meneses P., “Brecha de género en Matemáticas: El sesgo de las pruebas competitivas (evidencia para Chile),” no. 1140834, p. 119, 2016.
- [27] E. R. Ramirez-Saavedra, D. M. Espinosa-Sarmiento, and E. E. Millan-Rojas, “Estrategia para afrontar la deserción universitaria desde las tecnologías de la información y las comunicaciones,” *Revista Científica*, vol. 1, no. 24, p. 52, 2016.
- [28] M. Manco-Caycho, *Modelo predictivo para la identificación de patrones de la deserción estudiantil en la Universidad Nacional Tecnológica de Lima Sur (UNTELS)*. PhD thesis, 2016.
- [29] B. B. Mishra and S. Mishra, “Quality Improvements in Online Education System by Using Data Mining Techniques,” *Proceedings - 2nd International Conference on Data Science and Business Analytics, ICDSBA 2018*, vol. 5, pp. 532–536, 2018.
- [30] K. J. De Santos, A. G. Menezes, A. B. De Carvalho, and C. A. Montesco, “Supervised learning in the context of educational data mining to avoid university students dropout,” *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019*, vol. 2161-377X, pp. 207–208, 2019.
- [31] C. J. Sairy, “Minería de Datos para segmentación de clientes en la empresa tecnológica Master PC,” p. 184, 2015.
- [32] A. Jaramillo Zhingre, “Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje,” vol. 28, no. 1, pp. 64–90, 2015.
- [33] A. I. González Pineda, “Identificación de factores en la reprobación y deserción mediante técnicas de minería de datos en el área de la energía de la universidad nacional de Loja,” pp. 2–14, 2018.
- [34] R. O. María, “Determinación de perfiles profesionales mediante técnicas de minería de datos,” 2014.
-

- 
- [35] T. E. Abad Eras, “Aplicación De Minería De Datos Para Determinar Las Causas De La Violencia De Género Contra Las Mujeres En Ecuador,” p. 192, 2019.
- [36] J. D. Velásquez, “Una guía corta para escribir revisiones sistemáticas de literatura. Parte 1,” *DYNA (Colombia)*, vol. 82, no. 190, pp. 9–10, 2015.
- [37] J. D. Velásquez, “Una guía corta para escribir revisiones sistemáticas de literatura parte 2,” *DYNA (Colombia)*, vol. 82, no. 190, pp. 9–12, 2015.
- [38] J. D. Velásquez, “Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura. Part 3,” *DYNA (Colombia)*, vol. 82, no. 190, pp. 9–12, 2015.
-

# Lista de Acrónimos y Abreviaturas

<b>BGU</b>	Bachillerato General Unificado.
<b>EDM</b>	Minería de Datos Educativa.
<b>INEVAL</b>	Instituto Nacional de Evaluación Educativa.
<b>MD</b>	Minería de Datos.
<b>ODS</b>	Objetivos de Desarrollo Sostenible.
<b>PISA-D</b>	Programa internacional para la Evaluación de Estudiantes para el Desarrollo.
<b>SARA</b>	Sistema de Acompañamiento para el Rendimiento Académico.
<b>SENESCYT</b>	Secretaría de Educación Superior, Ciencia, Tecnología e Innovación.
<b>STEAM</b>	Ciencia, Tecnología, Ingeniería, Arte y Matemáticas..
<b>STEM</b>	Ciencia, Tecnología, Ingeniería y Matemáticas.
<b>TIC</b>	Tecnologías de la Información y la Comunicación.
<b>UIDE</b>	Universidad Internacional del Ecuador.
<b>UNESCO</b>	La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
<b>UNL</b>	Universidad Nacional de Loja.
<b>UTPL</b>	Universidad Técnica Particular de Loja.

# A. Anexo I

## A.1. Entrevistas a Estudiantes y Docentes CIS-CC

En esta sección se presenta una las entrevistas realizadas a docentes y estudiantes de las carreras de Ingeniería en Sistemas y Computación. De igual manera, las entrevistas de los estudiantes de bachillerato del Colegio "Bernardo Valdivieso". Cabe mencionar que se ha hecho uso de un repositorio en el cual están almacenadas todas las entrevistas ejecutadas.

**Click en el Enlace:**

[https://osf.io/9nmvb/?view\\_only=e0cedda2a98949a18ee3450487ba8601](https://osf.io/9nmvb/?view_only=e0cedda2a98949a18ee3450487ba8601)



Universidad  
Nacional  
de Loja



**Facultad de Energía, las Industrias y los Recursos Naturales no  
Renovables**

**CARRERA DE INGENIERÍA EN SISTEMAS - COMPUTACIÓN**  
**Entrevista a Docente de la Carrera de Ingeniería en Sistemas**

La siguiente entrevista se la realizó a la Ing. María del Cisne Ruilova Sánchez Docente de la CIS, en la que indicó "actualmente no existe estereotipos en las ingenierías, ya que estudiar una ingeniería es cuestión de preferencias, debido a que si una mujer desea ser parte de una ingeniería tiene la actitud, el interés y la posibilidad ya que ninguna universidad niega el acceso más bien buscan la equidad de género".

Considera que al someterse a un cambio como lo es del colegio a la universidad, se tiene claro la visión, la perspectiva de que lo que quiere en un futuro; añadiendo que el gusto de las matemáticas se da sin clasificar a nadie, puede ser a un hombre o una mujer, debido a que hay hombres que no les gusta las matemáticas y esto se puede dar en mujeres también.

Dentro del transcurso de preparación profesional, no existieron problemas con las matemáticas pero actualmente no las domina ya que para poder resolver un problema presente en la vida laboral en ámbitos específicos, se tendría que revisar, tomando en consideración que las matemáticas están en cualquier ámbito de la vida y ésta requiere siempre actitud y empeño desde etapas iniciales.

Nombre	Cédula	Cargo	Firma
María Ruilova S.	1103869770	Docente	

## B. Anexo II

### B.1. Revisión Sistemática de Literatura

Para examinar el estado de la cuestión sobre los factores académicos que influyen en la deserción universitaria aplicando técnicas de minería de datos, se busca obtener una base sólida de literatura científica, para ello se basará en las directrices presentes en [36–38] adaptándolas a las necesidades de TT.

Se contextualizará la RSL y delimitar planteando preguntas de investigación, las cuales permitirán que toda la RSL se oriente a la contestación de “¿*Cuáles son los factores más comunes presentes en la deserción estudiantil?*”, “¿*Las asignaturas relacionadas con las matemáticas influyen en la deserción académica?*”. y, ¿*Que técnicas de Minería de Datos se utilizan para determinar factores que influyen en la deserción universitaria?*”. Encaminados a la contestación a estas preguntas se plantean las siguientes estrategias de búsqueda de literatura:

- La literatura seleccionada debe estar dentro de los últimos 5 años de su publicación.
- Los documentos que serán utilizados en esta revisión bibliográfica serán solo los obtenidos en las fuentes bibliográficas detalladas a continuación:
  - **Google Scholar:** <https://scholar.google.com/>
  - **Scielo:** <http://www.scielo.org>
  - **Science Direct:** <https://www.sciencedirect.com/>
  - **Research Gate:** <https://www.researchgate.net/>
  - **IEEE** <https://www.ieee.org/>
- La búsqueda se basará en palabras específicas que den contestación a las preguntas de investigación, estas son: Minería de datos, deserción, abandono, matemáticas, software, dropout, student desertion, abandonment, mathematics, Data Mining, influence, factors,
- Para la búsqueda en las fuentes bibliográficas se usaran cadenas de búsqueda que contengan las palabras detalladas en el apartado anterior. La tabla B.1 muestra las cadenas de búsqueda que se usará para la RSL.

**Tabla B.1:** Cadenas de Búsqueda

Fuente Bibliográfica	Cadena de Búsqueda
Google Académico	("Desercion universitaria" minería de datos) AND (matematicas) AND (Años de publicación >2105)
Scielo	(Minería de Datos) AND Deserción AND ("Año de publicación >2015")
Science Direct	(Data Mining "Dropout")AND (Year of publication > 2015)
Research Gate	(deserción "minería de datos") AND (Año de publicación > 2015)
IEEE	Data Mining AND Mathematics AND Dropout

Buscando desechar la literatura que no ayude significativamente a contestar las preguntas de investigación, se descartará a la misma que no cumpla con:

- **Fecha de publicación:** Máximo 5 años atrás (2015)
- **Motores de búsqueda:** Google Scholar, Bibliotecas del Ecuador, Scielo, Science Direct, Research Gate, IEEE.
- **Idioma:** La literatura puede estar escrita en español o inglés.
- **Tipos de estudio:** Tesis, artículos científicos, libros.

Finalmente luego de establecer los lineamientos para ejecutar la RSL, se procede a realizar la búsqueda de literatura en las fuentes bibliográficas establecidas, aplicando las cadenas de búsqueda correspondientes a cada buscador. Al culminar con la búsqueda de literatura, aplicando los criterios de exclusión e inclusión se obtuvo un total de 14 trabajos relacionados que responderán las preguntas de investigación de la presente RSL, los trabajos relacionados se observan en la Tabla B.2:

**Tabla B.2:** Trabajos Relacionados

Título	Términos	Año de Publicación	Buscador
Minería de datos aplicada a estrategias para minimizar la deserción universitaria en carreras de Informática de la UNNOBA [13].	Deserción, Minería de datos	2019	Research Gate



**Tabla B.2:** Trabajos Relacionados

<b>Título</b>	<b>Términos</b>	<b>Año de Publicación</b>	<b>Buscador</b>
Estilos de aprendizaje y minería de datos: un estudio preliminar en el contexto universitario [19]	Deserción universitaria, Minería de datos	2018	Research Gate
Aplicación de técnicas de minería de datos para la identificación de patrones de deserción estudiantil como apoyo a las estrategias de SARA (sistema de acompañamiento para el rendimiento académico) [20].	Técnicas de Minería de Datos, abandono, deserción.	2018	Google Scholar
Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos [21].	Factores, Deserción. Data Mining.	2016	Google Scholar
Causas de Deserción en el Ingreso a la Universidad; un Estudio de Caso [6].	Deserción Universitaria, Abandono.	2017	Research Gate
Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos [22].	Deserción, Minería de Datos.	2015	Scielo
Predicción de deserción y éxito en estudiantes. Caso de estudio: ingeniería agroindustrial de la UNI Norte, Nicaragua, 2011-2015 [23].	Factores, Deserción. Minería de Datos.	2018	Research Gate

Tabla B.2: Trabajos Relacionados

Título	Términos	Año de Publicación	Buscador
Estrategia pedagógica con Aprendizaje basado en Juegos-GBL, para fomentar la motivación en el Área de Matemáticas en Ingeniería [24].	Minería de Datos, Deserción, Estrategias, Matemática	2019	Research Gate
Rendimiento académico, deserción y competencias matemáticas en ingeniería: caso Escuela de Ingeniería de Sistemas, UNSA [25].	Deserción, Rendimiento, Ingenierías.	2018	Research Gate
Brecha de género en matemáticas: el sesgo de las pruebas competitivas [26].	Matemáticas, Deserción	2016	Research Gate
Estrategia para afrontar la deserción universitaria desde las tecnologías de la información y las comunicaciones [27].	Deserción universitaria, Minería de Datos.	2016	Research Gate
Modelo Predictivo Para La Identificación De Patrones De La Deserción Estudiantil En La Universidad Nacional Tecnológica De Lima Sur (Untels) [28].	Predicción, deserción estudiantil.	2016	Google Scholar
Quality Improvements in Online Education System by Using Data Mining Techniques [29].	Dropout, mathematics, data mining	2018	IEEE
Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout [30].	Dropout, mathematics, data mining	2019	IEEE