



**UNIVERSIDAD NACIONAL DE LOJA**  
**ÁREA DE LA ENERGÍA, LAS INDUSTRIAS Y**  
**LOS RECURSOS NATURALES NO RENOVABLES**

**CARRERA DE INGENIERÍA EN SISTEMAS**

**“Análisis del proceso de Evaluación al Desempeño  
Docente, aplicando Técnicas de Minería de Datos.”**

*Tesis de Grado previa a la obtención del  
título de Ingeniero en Sistemas*

**AUTORA:**

*Janina Silvana Ortiz Pasaca.*

**DIRECTOR:**

*Ing. Luis Antonio Chamba Eras, Mg. Sc*

**LOJA - ECUADOR**  
**2014**

## **CERTIFICACIÓN**

Ing. Luis Antonio Chamba Eras, Mg. Sc

**DOCENTE DE LA CARRERA DE INGENIERÍA EN SISTEMAS**

### **CERTIFICA:**

Que la egresada Janina Silvana Ortiz Pasaca autora del presente trabajo de titulación, cuyo tema versa sobre **“ANÁLISIS DEL PROCESO DE EVALUACIÓN AL DESEMPEÑO DOCENTE, APLICANDO TÉCNICAS DE MINERÍA DE DATOS”**, ha sido dirigido, orientado y discutido bajo mi asesoramiento y reúne a satisfacción los requisitos exigidos en una investigación de este nivel por lo cual autorizo su presentación y sustentación.

Loja, Julio de 2014



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

## **AUTORÍA**

Yo **Janina Silvana Ortiz Pasaca** declaro ser autora del presente trabajo de tesis 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 tesis en el Repositorio Institucional- Biblioteca Virtual.

**Autor:** Janina Silvana Ortiz Pasaca

**Firma:**



**Cédula:** 1104902612

**Fecha:** 16 de Octubre de 2014

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

Yo, **Janina Silvana Ortiz Pasaca**, declaro ser autora de la tesis titulada: **Análisis del proceso de evaluación al desempeño docente, aplicando Técnicas de Minería de Datos**, como requisito para optar al grado de: **Ingeniera 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 las 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, dieciséis días del mes de octubre del dos mil catorce.

**Firma:**.....

**Autor:** Janina Silvana Ortiz Pasaca

**Cédula:** 1104902612

**Dirección:** Loja (Barrio Daniel Álvarez)      **Correo Electrónico:** jsortizp@unl.edu.ec

**Teléfono:** 2582447      **Celular:** 0990029705

**DATOS COMPLEMENTARIOS**

**Director de Tesis:** Ing. Luis Antonio Chamba Eras, Mg. Sc.

**Tribunal de Grado:** Ing. Luis Roberto Jácome Galarza, Mg. Sc.

Ing. Mario Andrés Palma Jaramillo, Mg. Sc.

Ing. Henry Patricio Paz Arias, Mg. Cc.



## **AGRADECIMIENTO**

En primer lugar quiero agradecer a Dios por darme la oportunidad de obtener un logro más en mi vida profesional y personal, igualmente extendiendo mi más sincero agradecimiento a la Institución en la cual me formé académicamente, Universidad Nacional de Loja, a los docentes y directivos que con sus enseñanzas y guías ayudan a que día a día, jóvenes puedan ver realizados sus metas y aspiraciones.

De manera especial agradezco al Ing. Luis Antonio Chamba Eras, por su guía, supervisión y ayuda en el presente Trabajo de Titulación.

**Janina Silvana Ortiz Pasaca.**

## **DEDICATORIA**

El presente trabajo está dedicado, a Dios de donde proviene toda sabiduría e inteligencia, por ser el principal soporte de mi vida, a mis padres Martha y Sergio por sus consejos, las enseñanzas aprendidas, su apoyo y amor entregado.

A la persona más noble con quien he podido contar en todo momento, a mi hermana Wendy por su amor y comprensión durante todo el desarrollo del presente trabajo, gracias por inspirarme a seguir y jamás rendirme, a levantarme y continuar.

**Janina Silvana Ortiz Pasaca.**

## **CESIÓN DE DERECHOS**

Janina Silvana Ortiz Pasaca autora principal del presente trabajo de titulación, autoriza a la Universidad Nacional de Loja, al Área de la Energía, las Industrias y los Recursos Naturales No Renovables y por ende a la Carrera de Ingeniería en Sistemas hacer uso del mismo en lo que estime sea conveniente.

## **a. Título**

“ANÁLISIS DEL PROCESO DE EVALUACIÓN AL DESEMPEÑO DOCENTE,  
APLICANDO TÉCNICAS DE MINERÍA DE DATOS.”

## **b. Resumen**

Los pasos agigantados que da la tecnología hoy en día, hace que cada vez exista mayor dependencia en el uso de sistemas de información y por ende almacenamiento de datos en todas las empresas o entidades públicas o privadas. Dentro de este gran conglomerado de generación de información, se encuentra el proceso de Evaluación de Desempeño Docente, que juega un papel muy importante en la calidad de la educación superior, ya que es un proceso, que busca la calidad y excelencia de cátedra.

El proceso de evaluación al desempeño docente en la actualidad, mide el desempeño en base a las puntuaciones obtenidas en el rendimiento de los docentes; que va en un rango de insatisfactorio, a un nivel muy satisfactorio, si bien se recoge la opinión del estudiante, este solo sirve para reforzar las preguntas ya establecidas dentro de un cuestionario, es decir no existe un proceso que evidencie la opinión de los alumnos en plan de mejora hacia el docente evaluado. Tampoco se permite conocer la evolución de cátedra que ha tenido un docente a través de los años, o los factores que influyen al momento que un estudiante evalúa a un docente.

En esta perspectiva se planteó el presente trabajo de titulación cuyo fin es realizar un análisis del proceso de evaluación de desempeño docente empleando técnicas de minería de datos para determinar el grado de influencia que tienen las notas de los estudiantes al momento de evaluar a un docente, el buen rendimiento o la existencia de criterios de interés dentro del proceso y la observación de la evolución de cátedra que ha tenido un docente en determinado tiempo.

A más de la aplicación de técnicas de minería de datos se hizo uso de herramientas de gestión de base de datos y herramientas de minería. En cuanto a la realización de cada proceso se basó en la metodología CRISP-DM que permitió culminar satisfactoriamente cada fase de minería. Los datos fueron obtenidos de la Unidad de Telecomunicaciones e Información (UTI) en físico y a través del Web Services, cabe destacar que los resultados del presente trabajo investigativo pueden ser utilizados para los fines convenientes ya sea en el apoyo a la toma de decisiones u optimización de procesos.



## **Summary**

The great steps that technology gives nowadays, makes that there is increasing reliance on the use of systems exist and therefore the storage of data in all businesses or public or private entities. Within this large conglomerate of information generation the process of Teaching Performance Assessment plays an important role in the quality of higher education. Because, it is a process that seeks quality and academic excellence.

The process of teaching performance assessment currently measures performance based on scores on the performance of teachers; going on a range of unsatisfactory at a very satisfactory level, although the student's opinion is collected, this only serves to reinforce the questions already established within a questionnaire, that is to say there is no process that evidences the views of students in improvement plan to the teacher evaluated. It also allowed the evolution of cathedra that had a teacher through the years, or the factors that influence a student when a teacher assesses.

In this perspective the present titling work whose aim is to analyze the process of evaluation of teaching performance using data mining techniques to determine the degree of influence that the notes have of the students when evaluating a teacher, good performance or the existence of criteria of interest in the process and observing the evolution of cathedra that a teacher have in time.

In addition to the application of data mining techniques made use of management tools and database mining tools. As for the performance of each process was based on the CRISP-DM methodology that allowed successfully completing each phase of mining. Data were obtained from the Telecommunications Unit and Information (UTI) in physical and through Web Services, note that the results of this research work can be used for suitable purposes or in support of decisions or optimization process.

# Índice de Contenidos

<b>a. Título.....</b>	<b>1</b>
<b>b. Resumen .....</b>	<b>2</b>
<b>Summary .....</b>	<b>3</b>
<b>Índice de Contenidos.....</b>	<b>4</b>
<b>Índice de Figuras .....</b>	<b>10</b>
<b>Índice de Tablas.....</b>	<b>13</b>
<b>c. Introducción.....</b>	<b>16</b>
<b>d. Revisión de Literatura .....</b>	<b>18</b>
1. CAPÍTULO I: Recopilación de Casos de Éxito en Fuentes Académicas, Revistas, Ponencias, Artículos Científicos, sobre la Aplicación de Minería de Datos en la Educación. ....	18
1.1 Caso De Éxito 1: Minería de Datos en la Educación .....	18
1.1.1 Introducción .....	18
1.1.2 Ventajas de la Minería de Datos dentro de la Investigación Educativa.....	19
1.1.3 Minería de Datos como solución .....	19
1.1.4 Estudios Realizados .....	20
1.1.5 Ejemplo de Estudio: Identificación de características de fracasos escolares en institutos .....	21
1.1.6 Recursos utilizados.....	23
1.2 Caso De Éxito 2: Minería de Datos en Encuestas de Profesores al Fin de Semestre de la Facultad de Ingeniería, UNAM. ....	24
1.2.1 Introducción .....	24
1.2.2 Objetivos del proyecto .....	24
1.2.3 Técnicas y Herramientas .....	24
1.2.4 Perspectivas del Proyecto.....	25
1.2.5 Resultados.....	26
1.2.6 Recursos utilizados.....	26
1.3 Caso De Éxito 3: Obtención de Patrones y Reglas en el Proceso Académico de La Universidad de Las Ciencias Informáticas Utilizando Técnicas de Minería de Datos. ....	27

1.3.1	Introducción .....	27
1.3.2	Técnicas y Herramientas .....	27
1.3.3	Compresión del Negocio.....	27
1.3.4	Resultados.....	28
1.3.5	Metodología .....	29
1.3.6	Recursos utilizados.....	29
2.	CAPÍTULO II: Recopilación De Técnicas De Minería De Datos. ....	30
2.1	Técnicas Predictivas o Supervisadas .....	31
2.1.1	Árboles de decisión.....	31
2.1.1.2	Estructura del árbol de decisiones .....	34
2.1.2	Redes Bayesianas .....	35
2.1.2.1	Naive Bayes.....	37
2.1.3	Redes neuronales.....	37
2.1.4	Regresión Logística .....	38
2.1.4.1	Ejemplo de Suscripción de un diario. ....	39
2.1.5	Algoritmos genéticos.....	40
2.1.6	Generación de Reglas .....	41
2.2	Técnicas Descriptivas o No Supervisadas.....	43
2.2.1	Agrupamiento o Clustering.....	43
2.2.2	Reglas de Asociación .....	44
2.2.2.1	Ejemplo de Reglas de Asociación.....	45
2.3	Algoritmos de Minería de Datos.....	46
3.	CAPÍTULO III: Herramientas Útiles Dentro De La Minería De Datos. ....	47
3.1	Knime .....	47
3.2	Weka .....	48
3.3	Rapidminer .....	49
	<b>e. Materiales y Métodos.....</b>	<b>51</b>
	<b>f. Resultados.....</b>	<b>59</b>
1.	FASE 1: Recopilación y Análisis de los Datos proporcionados con respecto al proceso de Evaluación al Desempeño Docente. ....	59
1.1	Realizar entrevistas con el responsable de la Unidad de Telecomunicaciones e Información (UTI) y con la coordinadora del proceso de evaluación de desempeño	

docente, para solicitar autorización en cuanto a manipulación de los datos y acceso a los mismos.....	59
1.2 Recopilación de los datos necesarios dentro del proceso de evaluación al desempeño docente, en el formato requerido. ....	60
1.3 Explorar y analizar los datos, para la toma de decisiones adecuadas en la generación de un Modelo de MD. ....	60
1.3.1 Exploración de la Información.....	60
1.3.1.1 Modelo de evaluación en la Universidad Nacional de Loja.....	60
1.3.1.2 Informantes de Calidad.....	61
1.3.1.3 Delimitación del Objeto a Evaluar .....	61
1.3.1.4 Códigos para los Ámbitos, Componentes y Criterios de Análisis .....	62
1.3.1.5 Ponderación.....	63
1.3.1.6 Valoración.....	64
1.3.2 Disposición de los datos. ....	64
1.4 Identificar la situación actual del proceso de Evaluación de Desempeño Docente.....	80
2. FASE 2: Analizar las técnicas de Minería de Datos (MD) factibles para el proceso de Evaluación al Desempeño Docente .....	89
2.1 Búsqueda de información en fuentes académicas, revistas, tesis doctorales y artículos científicos sobre las diversas técnicas de MD.....	89
2.2 Investigar y analizar herramientas útiles en el proceso, y aplicación de técnicas de MD.....	92
2.2.1 Herramientas útiles en el proceso de Minería de Datos .....	92
2.2.1.1 Evaluación final.....	93
2.2.2 Herramientas, útiles para gestión de Bases de Datos.....	94
2.2.3 Herramientas, de apoyo al proceso de Minería de Datos.....	96
2.3 Selección de las herramientas y técnicas para el análisis de evaluación docente.....	98
2.3.1 Selección de Herramientas .....	98
2.3.2 Selección de Técnicas .....	99
3. FASE 3: Proceso De Minería De Datos .....	100
3.1 Procesos y Tratamiento de los datos en la elaboración de un modelo de minería de datos.....	100
3.1.1 Comprensión del Negocio o Problema.....	100
3.1.1.1 Determinar los objetivos del negocio.....	100

3.1.1.2 Evaluación de la Situación .....	101
3.1.1.2.1 Recursos Disponibles .....	102
3.1.1.2.1.1 Recursos Humanos.....	102
3.1.1.2.1.2 Recursos Hardware: .....	103
3.1.1.2.1.3 Recursos Software:.....	103
3.1.1.2.1.4 Datos: .....	103
3.1.1.2.1.5 Materiales: .....	104
3.1.1.2.2 Riesgos y contingencias. ....	104
3.1.1.2.3 Terminología.....	105
3.1.1.2.4 Presupuesto.....	107
3.1.1.2.4.1 Costes de Talento Humano.....	107
3.1.1.2.4.2 Costes Recursos Hardware .....	108
3.1.1.2.4.3 Costes Recursos Software.....	109
3.1.1.2.4.4 Materiales y Servicios .....	109
3.1.1.2.4.5 Coste Total.....	111
3.1.1.2.5 Cronograma del Proyecto .....	111
3.1.1.3 Determinación de metas de la minería de datos .....	113
3.1.1.4 Plan del Proyecto.....	114
3.1.2 Comprensión de los Datos.....	116
3.1.2.1 Recolección Inicial de los Datos.....	116
3.1.2.2 Descripción de los datos. ....	118
3.1.2.3 Exploración de los datos .....	130
3.1.2.4 Verificación de la calidad de los datos. ....	132
3.1.3 Preparación de los Datos.....	132
3.1.3.1 Selección de los Datos.....	133
3.1.3.2 Limpieza de los Datos.....	133
3.1.3.3 Construcción de Datos.....	134
3.1.3.4 Integración de datos.....	141
3.1.4 Modelado.....	144
3.1.4.1 Selección de la técnica de modelado.....	144
3.1.4.1.1 Salidas del Modelo.....	146



3.1.4.2 Generación del diseño de pruebas.....	146
3.1.4.3 Construcción de Modelo .....	147
3.1.4.3.1 Criterios de Interés existentes en el proceso de evaluación docente. ....	147
3.1.4.3.1.1 Clasificación mediante ID3.....	148
3.1.4.3.1.2 Clasificación mediante Árbol CHAID .....	150
3.1.4.3.1.3 Clasificación mediante Árbol C4.5.....	151
3.1.4.3.1.4 Generación de Reglas mediante el algoritmo W-PART .....	153
3.1.4.3.1.5 Generación de Reglas mediante el algoritmo W-NNge .....	155
3.1.4.3.1.6 Modelos W-Ridor, W-ConjunctiveRule .....	156
3.1.4.3.1.7 Obtención de Opiniones:.....	157
3.1.4.3.2 Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos. ....	158
3.1.4.3.2.1 Clasificación Naive Bayes .....	158
3.1.4.3.2.2 Clasificación mediante ID3.....	160
3.1.4.3.2.3 Clasificación mediante W-RETree.....	161
3.1.4.3.2.4 Reglas mediante W-NNge.....	163
3.1.4.3.3 Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente. ....	165
3.1.4.3.3.1 Regresión logística.....	165
3.1.4.3.3.2 Clasificación mediante ID3.....	168
3.1.4.3.3.3 Clasificación mediante C4.5.....	169
4. FASE 4: Evaluación del modelo generado e interpretación de Resultados .....	172
4.1 Evaluación Modelo .....	172
4.1.2 Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos. ....	175
4.1.3 Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente. ....	178
4.2 Validación del Modelo y Análisis de Resultados. ....	180
4.2.1 Evaluación y Validación .....	180
4.2.1.1 Evaluación de Resultados.....	181
4.2.1.1.1 Resultados de Criterios de Interés existentes en el proceso de evaluación docente.....	181

4.2.1.1.2 Resultados Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos.....	195
4.2.1.1.3 Resultados en el Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente. ....	200
4.2.1.2 Evaluación de los resultados de la minería de datos con respecto a Factores Críticos del Éxito.....	205
5. FASE 5: Culminación de la redacción de la memoria final del Proyecto Fin de Carrera. ....	207
5.1 Redacción de un artículo científico basado en el estándar IEEE.....	207
<b>g. Discusión</b> .....	<b>208</b>
1. Desarrollo de la propuesta alternativa.....	208
2. Valoración Técnica Económica Ambiental .....	211
<b>h. Conclusiones</b> .....	<b>214</b>
<b>i. Recomendaciones</b> .....	<b>215</b>
<b>j. Bibliografía</b> .....	<b>216</b>
<b>k. Anexos</b> .....	<b>227</b>
ANEXO 1: Acuerdo de Confidencialidad .....	227
ANEXO 2: Autorización Acceso a los Datos de Evaluación de Desempeño Docente ....	229
ANEXO 3: Autorización Acceso a los Datos del Web Services .....	230
ANEXO 4: Cuestionario de Evaluación Docente y Satisfacción Estudiantil. ....	231
ANEXO 5: Datos Recopilados .....	236
ANEXO 6: Herramientas útiles en el proceso de Minería de Datos.....	238
ANEXO 7: Procesos y pasos de Minería de Datos operadores en Rapid Miner.....	246
ANEXO 8: Análisis General: Escenario Real .....	274
ANEXO 9: Nuevos Datos Evaluación De Desempeño Docente .....	286
ANEXO 10: Informe Ejecutivo.....	287
ANEXO 11: Análisis Encuesta de Evaluación de Desempeño Docente .....	303
ANEXO 12: Certificación Traducción Summary .....	311

## Índice de Figuras

Figura 1: Conjunto de Clústers .....	21
Figura 2: Reglas de asociación.....	21
Figura 3: Fragmento modelo árbol de decisión.....	25
Figura 4: Técnica árbol de decisión .....	32
Figura 5: Estructura del árbol de decisión.....	35
Figura 6: Ejemplo de una red bayesiana.....	36
Figura 7: Técnica red neuronal.....	38
Figura 8: Técnica algoritmos genéticos.....	41
Figura 9: Técnica Clustering. ....	43
Figura 10: Encuesta realizada por la KDnuggets Metodologías utilizadas en MD.....	53
Figura 11: Modelo de proceso CRISP-DM.....	54
Figura 12: Categorías e información dentro del Web Services. ....	67
Figura 13: Estudiantes del AEIRNNR por Carrera y Módulo. ....	69
Figura 14: Estudiantes del AEIRNNR por Carrera y Módulo. ....	70
Figura 15: Estudiantes del AEIRNNR por Carrera y Módulo. ....	71
Figura 16: Estudiantes del AEIRNNR por Carrera y Módulo .....	71
Figura 17: Número estudiantes- Periodos Académicos.....	75
Figura 18: Número de estudiantes por periodo académico. ....	75
Figura 19: Docentes por periodos de evaluación.....	78
Figura 20: Docentes por periodos de evaluación.....	79
Figura 21: Formatos de archivos por cada herramienta.....	95
Figura 22: Cronograma de Proyecto.....	112
Figura 23: BD en el proceso de EDD.....	117
Figura 24: Estructura de la Base de Datos EDD.....	129
Figura 25: Diagrama de la base de datos final.....	143
Figura 26: Matriz de confusión árbol ID3. ....	149
Figura 27: Árbol de clasificación ID3.....	149
Figura 28: Matriz de confusión CHAID .....	151
Figura 29: Árbol de clasificación CHAID. ....	151
Figura 30: Matriz de confusión C4.5 .....	152
Figura 31: Árbol de clasificación C4.5 .....	153
Figura 32: Matriz de confusión W-PART .....	154
Figura 33: Reglas mediante W-PART.....	154
Figura 34: Matriz de confusión W-NNge .....	155
Figura 35: Reglas mediante W-NNge .....	156
Figura 36: Reglas mediante W-Ridor.....	157
Figura 37: Reglas mediante W-ConjunctiveRule .....	157

Figura 38: Opiniones relevantes, Many eyes .....	158
Figura 39: Matriz de confusión Naive Bayes .....	159
Figura 40: Clasificación Naive Bayes.....	159
Figura 41: Matriz de confusión árbol ID3. ....	161
Figura 42: Árbol de clasificación ID3.....	161
Figura 43: Matriz de confusión W-RETree.....	162
Figura 44: Algoritmo W-REPTree.....	163
Figura 45: Matriz de confusión W-NNge .....	164
Figura 46: Algoritmo W-NNge .....	164
Figura 47: Generación de reglas W-Jrip.....	165
Figura 48: Matriz de confusión algoritmo de regresión logística .....	166
Figura 49: Modelo regresión logística.....	167
Figura 50: Descripción y análisis gráfico.....	167
Figura 51: Matriz de confusión algoritmo ID3.....	168
Figura 52: Árbol de clasificación ID3.....	169
Figura 53: Matriz de confusión C4.5 .....	170
Figura 54: Árbol de decisión generado C4.5 .....	171
Figura 55: Resultados obtenidos por cada algoritmo Criterios Interés.....	174
Figura 56: Resultados obtenidos por cada clase Criterios Interés.....	175
Figura 57: Resultados obtenidos por cada algoritmo- Evolución Cátedra.....	176
Figura 58: Resultados obtenidos por cada clase, Evolución Cátedra .....	177
Figura 59: Resultados obtenidos por cada algoritmo, Notas. ....	179
Figura 60: Resultados obtenidos por cada clase, Notas. ....	180
Figura 61: Criterios de interés y Buen rendimiento mediante algoritmo ID3. ....	181
Figura 62: Descripción y análisis gráfico.....	182
Figura 63: Docentes clasificados mediante el algoritmo ID3. ....	183
Figura 64: Criterios de interés y Buen rendimiento mediante algoritmo CHAID.....	183
Figura 65: Docentes clasificados mediante el algoritmo CHAID. ....	184
Figura 66: Porcentaje Opiniones Estudiantes por carrera en el proceso de Evaluación..	189
Figura 67: Opiniones Estudiantes dentro del proceso de Evaluación .....	189
Figura 68: Palabras recurrentes dentro de las opiniones-Criterios Interés.....	190
Figura 69: Palabras con mayor ocurrencia: Enseñanza .....	191
Figura 70: Opiniones Estudiantes dentro del proceso de Evaluación .....	191
Figura 71: Palabras con mayor ocurrencia: Mejorar .....	192
Figura 72: Opiniones Estudiantes dentro del proceso de Evaluación .....	192
Figura 73: Palabras con mayor ocurrencia: Mejorar .....	193
Figura 74: Opiniones Estudiantes: Plan de Mejora .....	193
Figura 75: Incidencia de palabras por carreras.....	194
Figura 76: Opiniones Estudiantes dentro del proceso de Evaluación .....	194
Figura 77: Modelo Naive Bayes.....	196

Figura 78: Docentes clasificados mediante el algoritmo Naive Bayes. ....	196
Figura 79: Evolución de cátedra ID3.....	197
Figura 80: Docentes clasificados mediante el algoritmo ID3. ....	197
Figura 81: Predicción de calificaciones en el proceso de evaluación de desempeño. ....	201
Figura 82: Número de estudiantes que califican por categoría. ....	202
Figura 83: Clasificación calificación docente periodo académico 1.....	202
Figura 84: Clasificación calificación docente periodo académico 2.....	203
Figura 85: Clasificación calificación docente periodo académico 3.....	203
Figura 86: Clasificación modelo C4.5 .....	204
Figura 87: Clasificación calificación docente por categorías. ....	204



## Índice de Tablas

TABLA I: Validación Cruzada.....	23
TABLA II: Técnicas, Herramientas, Algoritmos .....	23
TABLA III: Técnicas, Algoritmos Y Herramientas .....	27
TABLA IV: Técnicas, Herramientas.....	29
TABLA V: Técnicas De Minería De Datos.....	31
TABLA VI: Datos Empleados En Regresión Lineal .....	40
TABLA VII: Resultados De Regresión Logística .....	40
TABLA VIII: Transacciones Para Producir Reglas De Asociación .....	45
TABLA IX: Metodologías Md .....	52
TABLA X: Informantes De Calidad.....	61
TABLA XI: Ámbito Y Componentes Evaluados .....	61
TABLA XII: Códigos Para Componentes Evaluados.....	62
TABLA XIII: Escala De Valoración Cuantitativa. ....	63
TABLA XIV: Peso Cuantitativo Para Cada Componente.....	63
TABLA XV: Peso Cuantitativo Por Informantes .....	64
TABLA XVI: Equivalente Quali-Cuantitativo Del Proceso Evaluativo. ....	64
TABLA XVII: Datos Correspondientes Carreras Asignaturas. ....	65
TABLA XVIII: Datos Correspondientes Estudiantes Docentes .....	65
TABLA XIX: Datos Correspondientes Cuestionario Evaluación.....	66
TABLA XX: Descripción De Los Servicios Por Categorías. ....	67
TABLA XXI: Periodos Académicos De Edd.....	68
TABLA XXII: Periodos De Evaluación .....	68
TABLA XXIII: Carreras Aeirnr. ....	73
TABLA XXIV: Id Periodos Académicos Y Periodos De Evaluación.....	73
TABLA XXV: Id Ofertas Académicas.....	73
TABLA XXVI: Relación Entre Tablas De La Edd.....	74
TABLA XXVII: Nomenclatura De Componentes.....	74
TABLA XXVIII: Total Docentes .....	76
TABLA XXIX: Total Estudiantes Y Docentes.....	76
TABLA XXX: Total Evaluaciones.....	77
TABLA XXXI: Docentes Por Periodo Académico.....	77
TABLA XXXII: Docentes Por Carrera Y Periodo Académico .....	79
TABLA XXXIII: Recopilación Técnicas De Minería De Datos .....	89
TABLA XXXIV: Características De Herramientas De Minería De Datos.....	93
TABLA XXXV: Características Principales Herramientas Md - Evaluación Final .....	93
TABLA XXXVI: Características De Herramientas Para Administrar Bd.....	95
TABLA XXXVII: Herramientas De Apoyo Al Proceso De Minería .....	96
TABLA XXXVIII: Selección Técnicas De Minería De Datos .....	99

TABLA XXXIX: Riesgos y Contingencias Del Proyecto.....	104
TABLA XL: Costo Por Hora De Integrantes De Proyecto.....	107
TABLA XLI: Actividades Del Proyecto Y Duración. ....	107
TABLA XLII: Asignación De Actividades Por Integrante y Número De Horas. ....	108
TABLA XLIII: Recopilación De Horas Y Costes Por Rol Del Personal.....	108
TABLA XLIV: Costes Del Hardware.....	109
TABLA XLV: Costes Del Software.....	109
TABLA XLVI: Costes Materiales Y Servicios.....	110
TABLA XLVII: Gastos Adicionales.....	110
TABLA XLVIII: Coste Total Del Proyecto .....	111
TABLA XLIX: Plan Del Proyecto.....	115
TABLA L: Descripción De Los Servicios Por Categorías. ....	117
TABLA LI: Tabla Carreras Y Asignaturas.....	118
TABLA LII: Tablas Estudiantes Y Docentes .....	119
TABLA LIII: Tabla Cuestionario Y Evaluación.....	119
TABLA LIV: Métodos Utilizados Para Obtener Información De Categoría Académica. ...	120
TABLA LV: Estructura De Tabla App_Areasga.....	122
TABLA LVI: Estructura De Tabla Carrera. ....	122
TABLA LVII: Estructura De Tabla App_Asignatura. ....	122
TABLA LVIII: Estructura De Tabla App_Asignaturadocente .....	122
TABLA LIX: Estructura De Tabla App_Estudiante_Asig_Doc.....	123
TABLA LX: Estructura De Tabla App_Estudianteperiodoacademico .....	123
TABLA LXI: Estructura De Tabla Estudiante_Evaluacion .....	123
TABLA LXII: Estructura De Tabla App_Docenteperiodoacademico .....	124
TABLA LXIII: Estructura De Tabla Docente_Evaluacion .....	124
TABLA LXIV: Estructura De Tabla App_Cuestionario.....	124
TABLA LXV: Estructura De Tabla App_Seccion.....	125
TABLA LXVI: Estructura De Tabla App_Pregunta.....	125
TABLA LXVII: Estructura De Tabla App_Tipopregunta .....	125
TABLA LXVIII: Estructura De Tabla App_Itempregunta .....	125
TABLA LXIX: Estructura De Tabla App_Contestacion.....	126
TABLA LXX: Estructura De Tabla App_Evaluacion.....	126
TABLA LXXI: Estructura De Tabla App_Periodoevaluacion .....	126
TABLA LXXII: Estructura De Tabla App_Periodoevaluacion_Areassga .....	127
TABLA LXXIII: Estructura De Tabla App_Periodoacademico.....	127
TABLA LXXIV: Estructura De Tabla App_Periodoacademico_Ofertasacademicassga ...	127
TABLA LXXV: Estructura De Tabla App_Ofertaacademicassga.....	127
TABLA LXXVI: Estructura De Tabla App_Tipoinformante.....	127
TABLA LXXVII: estructura de tabla app_usuario.....	128
TABLA LXXVIII: Estructura De Tabla Auth_User.....	128

TABLA LXXIX: Estructura De Tabla Nota_Unidad_Estudiente .....	128
TABLA LXXX: Número De Estudiantes Por Carreras.....	130
TABLA LXXXI: total docentes.....	131
TABLA LXXXII: Total Evaluaciones.....	131
TABLA LXXXIII: Limpieza De Datos .....	134
TABLA LXXXIV: Atributos En La Estructura De Minería De Datos.....	135
TABLA LXXXV: Estructura 1 .....	138
TABLA LXXXVI: Estructura 2,3.....	139
TABLA LXXXVII: Estructura 4 .....	140
TABLA LXXXVIII: Rendimiento, Modelo, Algoritmo Id3.....	148
TABLA LXXXIX: Rendimiento, Modelo Chaid .....	150
TABLA XC: Rendimiento, Modelo C4.5.....	152
TABLA XCI: Rendimiento, Modelo W-Part.....	154
TABLA XCII: RENDIMIENTO, MODELO W-Nnge.....	155
TABLA XCIII: Rendimiento, Modelo, Algoritmo Naive Bayes .....	159
TABLA XCIV: Rendimiento, Modelo, Algoritmo Id3.....	160
TABLA XCV: Rendimiento, Modelo, Algoritmo Basado En W-Retree.....	162
TABLA XCVI: Rendimiento, Modelo, Algoritmo Basado En W-Nnge .....	163
TABLA XCVII: Rendimiento, Modelo De Regresión Logística.....	166
TABLA XCVIII: Rendimiento, Modelo, Algoritmo Id3.....	168
TABLA XCIX: Rendimiento, Modelo, C4.5.....	170
TABLA C: Evaluación De Modelos Generados En Criterios De Interés.....	172
TABLA CI: Evaluación De Modelos Generados Evolución De Cátedra.....	176
TABLA CII: Evaluación De Modelos Generados En Influencia De Notas.....	178
TABLA CIII: Asignaturas Que Generan Criterios De Interés.....	184
TABLA CIV: Asignaturas Buen Rendimiento.....	186
TABLA CV: Talento Humano .....	212
TABLA CVI: Recursos Hardware.....	212
TABLA CVII: Recursos Software.....	213
TABLA CVIII: materiales y servicios.....	213
TABLA CIX: Coste Total Del Proyecto.....	213

## **c. Introducción**

Se sabe que las instituciones tanto públicas o privadas generan y almacenan gran cantidad de información debido al avance tecnológico y uso de las mismas. Por ende se ha catalogado a la información como el activo más importante de una empresa o institución y la producción de la misma abarca la generación de estrategias, incremento de ganancias, automatización de procesos, pero sobre todo apoyo a la toma de decisiones.

Existen diversos procesos dentro de un ambiente educativo que producen y almacenan gran cantidad de información, entre ellos se encuentra el proceso de Evaluación del Desempeño Docente (EDD) [1], como proceso en la calidad de la educación superior [2]. Se puede referir a la evaluación del desempeño docente, como un conjunto de acciones organizadas en base a las intervenciones e interacciones pedagógicas, entre el conocimiento científico y el conocimiento académico, lo cual permite impulsar acciones didácticas pedagógicas que fortalezcan los procesos de aprendizaje de los estudiantes, y el mejoramiento de la formación inicial docente, y a su vez su desarrollo profesional [3-7].

La información que proviene de dicho proceso constituye un instrumento útil, bajo determinadas circunstancias. Pero lo cierto es que, institucionalmente, no tienen mayor incidencia en la gestión universitaria, debido a la falta de procesos apropiados que filtren o procesen este tipo de información [8,9], ya que generalmente es demasiada, dando como consecuencia una carencia de elaboración de estrategias de mejoras en función de las necesidades que surjan de los mismos [5].

Herramientas o técnicas que proporcionen un análisis en este tipo de procesos educativos son importantes [10-17] para conocer en este caso, los factores y comportamientos en la evaluación de desempeño docente.

Por tal motivo y con un enfoque a lo anteriormente descrito, el presente trabajo de titulación se enmarca, a la aplicación de técnicas de minería de datos para la generación de modelos que permitan la obtención de información o comportamientos dentro del proceso de evaluación.

Dichos modelos permitirán conocer el grado de influencia que tienen las notas de los estudiantes en la calificación de desempeño docente, los criterios de interés que pueden existir hacia un docente al momento de ser evaluado y la evolución de cátedra que ha tenido un docente a través de los años o periodos académicos.

El desarrollo del trabajo de titulación empieza con un análisis de las técnicas de minería de datos (MD), luego se realizó un análisis de los datos proporcionados por la UTI, posteriormente se recolectó y analizó casos de éxito que involucre la minería en el ámbito educativo, para luego dar paso al desarrollo de las fases de minería [18] y finalmente generar modelos que permitan evidenciar los resultados esperados.

Los resultados alcanzados por el presente trabajo de titulación se replicaron a la Comunidad Científica esto con el objetivo de contribuir a las personas que lo estimen conveniente.

Se debe destacar que la realización del presente trabajo investigativo puso en práctica los conocimientos adquiridos a lo largo de la vida académica, siendo a la vez dichos conocimientos útiles en el ámbito profesional.

## **d. Revisión de Literatura**

### **1. CAPÍTULO I: RECOPIACIÓN DE CASOS DE ÉXITO EN FUENTES ACADÉMICAS, REVISTAS, PONENCIAS, ARTÍCULOS CIENTÍFICOS, SOBRE LA APLICACIÓN DE MINERÍA DE DATOS EN LA EDUCACIÓN.**

La información siendo el activo más importante de una institución debe ser tratada mediante procesos y herramientas que permitan la generación de estrategias, incremento de ganancias, automatización de procesos, pero sobre todo apoyo a la toma de decisiones. Es así que entidades de educación superior generan gran cantidad de información que puede ser analizada mediante el uso de técnicas de minería de datos tanto para el descubriendo de conocimiento como para la descripción de factores o comportamientos no muy evidentes en primera instancia.

A continuación se presenta una recopilación de casos de éxito en el ámbito educativo útiles en el desarrollo del presente trabajo de titulación, ya que los mismos dieron pautas para su realización [12], [19-21].

#### **1.1 CASO DE ÉXITO 1: Minería de Datos en la Educación**

El caso de éxito presentado a continuación, trata de la aplicación de minería de datos en la educación; como solución a diversos problemas inmersos dentro de este escenario.

##### **1.1.1 Introducción**

La minería de datos, es el campo de la informática que permite descubrir información nueva y potencialmente útil de grandes cantidades de datos. Se ha empleado en numerosos campos, incluyendo desde los ya conocidos casos de cesta de la compra hasta la bioinformática o investigaciones contra el terrorismo. Recientemente, se ha incrementado el interés en utilizar la minería de datos en el estudio educacional, centrándose en el desarrollo de métodos de descubrimiento que utilicen los datos de plataformas educacionales y en el uso de esos métodos para comprender mejor a los

estudiantes y el entorno en el que aprenden. Métodos psicométricos suelen ser integrados con métodos de aprendizaje máquina y textos de minería de datos para lograr los objetivos. Por ejemplo, obteniendo datos sobre cómo los estudiantes eligen utilizar el software educacional, puede ser realmente útil considerar datos a distintos niveles sobre las pulsaciones de teclas, nivel de respuestas del alumno [12].

### **1.1.2 Ventajas de la Minería de Datos dentro de la Investigación Educativa**

La minería de datos educacionales 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. En particular, la creación de repositorios públicos de datos educacionales ha creado una base que hace posible la minería de datos educacionales. En particular, los datos de estos repositorios son totalmente válidos (ya que son datos reales sobre el rendimiento y aprendizaje de estudiantes reales, en ambientes educacionales, tomados en tareas de aprendizaje), y cada vez más fácilmente accesibles para comenzar una investigación [12].

### **1.1.3 Minería de Datos como solución**

La minería de datos orientada a la educación permite predecir cualquier tipo de factor o característica de un caso, fenómeno o situación. De esta forma, utilizando las técnicas que ofrece la minería, se puede predecir, con un porcentaje muy alto de credibilidad, por ejemplo la probabilidad de desertar de cualquier alumno con la ventaja de que se puede pronosticar en los primeros cuatrimestres.

La minería de datos en la educación ha venido utilizándose considerablemente en los últimos años.

Sobre este tema se han hecho algunas investigaciones muy similares, tal es el caso de la Universidad de La Sabana en Colombia, en donde el objetivo era seleccionar, de una base de datos de estudiantes, los atributos que tuvieran mayor incidencia en la deserción de la Universidad en los últimos cuatro años, para este estudio se utilizó una técnica de minería de datos llamada Rough Sets.

Otro estudio reciente sobre las correlaciones existentes en este tema es el de Wayman en 2001. Se trata de un estudio muy técnico que toma muchas medidas de los estudiantes, a

través de un modelo de regresión logística y de imputación múltiple. Gracias a este estudio se encontró un conjunto muy potente de predictores basados en la recompensa del estudio, el nivel socio-económico, y la edad [12, 19].

#### **1.1.4 Estudios Realizados**

Se han realizado estudios sobre minería de datos en sistemas educativos basados en tecnologías web, como educación a distancia o asistida por computadora, en la Universidad Leonardo Da Vinci en Francia y la Universidad de Sydney en Australia, entre otras universidades, mostraron cómo utilizar los algoritmos de minería de datos para descubrir conocimiento pedagógico relevante que se almacenaba en bases de datos. Estos descubrimientos ayudaron, tanto a docentes como administrativos y directivos a entender el aprendizaje de sus estudiantes y ofrecer sus enseñanzas de una mejor manera hacia ellos [12, 19].

#### **Técnicas Utilizadas:**

- 1) **Agrupamiento:** En agrupamiento, el objetivo es encontrar puntos de datos que se agrupen de manera natural, repartiendo el conjunto original de datos en un conjunto de 'clústeres' (Ver Figura1 [12]). El agrupamiento es particularmente útil en casos donde las categorías más comunes de los datos no son conocidas. Si un conjunto de clústeres es óptimo, en cada categoría, cada punto será más similar a los puntos pertenecientes a su clúster que a puntos pertenecientes a otros grupos. Los clústeres pueden ser creados con distinta granularidad: por ejemplo, las escuelas podrían ser agrupadas para investigar similitudes y diferencias entre ellas, los estudiantes podrían también agruparse por el mismo motivo, o incluso podrían agruparse las acciones de los estudiantes para investigar patrones de comportamiento. Los algoritmos de agrupamiento pueden comenzar sin hipótesis previas sobre los grupos de datos (como el algoritmo k-means con inicio aleatorio), o empezar desde una hipótesis específica, posiblemente generada en estudios previos con un conjunto de datos distinto [12].



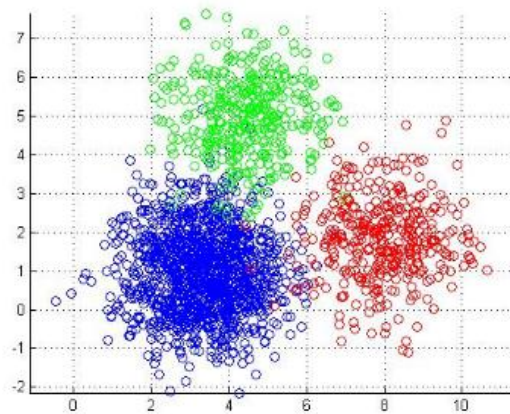


Figura 1: Conjunto de Clústers

2) **Minería de Relaciones:** En la minería de relaciones, el objetivo es descubrir relaciones entre variables en un conjunto de datos con un gran número de variables. Una forma de realizar esto es localizando las variables más fuertemente relacionadas con una única variable de interés, o también mediante el descubrimiento de las relaciones más fuertes entre dos variables. En la minería de reglas de asociación, el objetivo es encontrar reglas ‘si X entonces Y’ (Ver Figura2 [12]) de manera que si se encuentra un conjunto de variables, otra variable tendrá habitualmente un determinado valor. Por ejemplo, una regla podría ser:

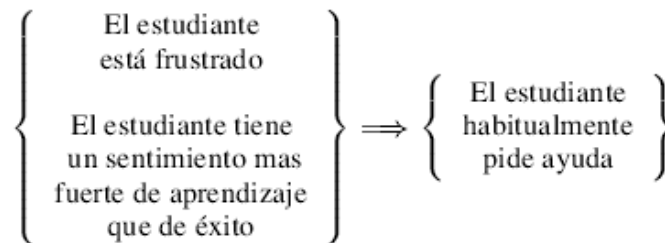


Figura 2: Reglas de asociación

En la minería de regresiones, el objetivo es encontrar correlaciones lineales o logísticas entre variables [12].

### 1.1.5 Ejemplo de Estudio: Identificación de características de fracasos escolares en institutos

En este ejemplo se toma una estructura generada en base a las siguientes variables:

- Grupo (Abandono escolar/Graduado escolar)

- Edad en años
- Sexo
- Grupo étnico
- Status socio-económico
- N° de infracciones disciplinarias de nivel 1
- N° de infracciones disciplinarias de nivel 2
- N° de infracciones disciplinarias de nivel 3
- N° de infracciones disciplinarias de nivel 4
- N° de clases avanzadas a las que ha asistido
- N° de clases de Matemáticas a las que ha asistido
- N° de clases de Ciencias a las que ha asistido
- Faltas justificadas
- Faltas sin justificar
- Nota media
- Nivel “CSAP Reading Proficiency” (entre cuatro niveles)
- Nivel “CSAP Writing Proficiency” (entre cuatro niveles)
- Nivel “CSAP Math Proficiency” (entre cuatro niveles)

Algunas de estas variables se estimaron a partir de otros datos. Por ejemplo, el nivel socio-económico, fue estimado según si el colegio era gratuito o no. El nivel de infracción disciplinaria va de menor a mayor, siendo 1 el primero, y 4 el último. Como resultado de una ejecución preliminar usando todas estas variables, algunas de ellas fueron recodificadas con el fin de reducir el número de categorías y hacer los árboles resultantes más interpretables. Se recodificaron “edad”, “grupo étnico” y “nota media” y el número de clases de matemáticas, ciencias, y avanzadas. También, como el nivel “1: Insatisfactorio” y “4: Avanzado” ocurrían relativamente poco, se simplificaron las variables del CSAP con valores dicotómicos (1: Aprobado, 0: Suspenso) [12].

Aplicando diversas técnicas de minería se obtuvieron modelos los mismos que fueron evaluados con la matriz de confusión, esto con el fin de comprobar el rendimiento y comportamiento del modelo (Ver Tabla I [12]).

TABLA I: VALIDACIÓN CRUZADA

Matriz de clasificación errónea				
Categoría Actual				
		Fracaso	Graduado	Total
Categoría Estimada	Fracaso	562	116	678
	Graduado	65	403	468
	Total	627	519	1146
Riesgo Estimado		Validación Cruzada		
0.158		0.172		

Los resultados obtenidos en el presente ejemplo de estudio relacionan el fracaso escolar, con el rendimiento académico, las notas individuales de los estudiantes dando como conclusiones del modelo que, para los estudiantes de la nota más baja, se les clasifica según la edad, de tal manera que si tiene mala nota y mucha edad tiene más probabilidad de fracaso que si tiene mala nota pero es más joven.

### 1.1.6 Recursos utilizados

En base al caso de éxito y dentro de la educación se pueden utilizar técnicas (Ver Tabla II [12]).herramientas y algoritmos descritos a continuación:

TABLA II: TÉCNICAS, HERRAMIENTAS, ALGORITMOS

Técnicas	Algoritmos	Descripción
Agrupamiento-Clúster	K-means	Encuentra puntos de datos que se agrupen de manera natural, sin necesidad de que exista una etiqueta o label a predecir.
Árbol de Decisión	CHAID	Examina las relaciones entre muchas variables de predicción categóricas, ordinales, o continuas, y la variable objeto de estudio, como herramientas de apoyo fueron Answer Tree, SPSS.
Generación de Reglas	Reglas asociación	Descubre relaciones entre variables en un conjunto de datos.
Regresión	<ul style="list-style-type: none"> <li>▪ Regresión logística</li> <li>▪ Regresión Lineal</li> </ul>	Explica el comportamiento de una variable dependiente discreta, $Y$ , en función de una o más variables independientes

## **1.2 CASO DE ÉXITO 2: Minería de Datos en Encuestas de Profesores al Fin de Semestre de la Facultad de Ingeniería, UNAM.**

A continuación se presenta el caso de éxito de la minería de datos aplicada a encuestas de profesores de la Universidad Nacional Autónoma de México.

### **1.2.1 Introducción**

El presente caso de éxito se enfoca en el análisis del desempeño docente, en base a una encuesta de profesores realizada cada fin de semestre a los alumnos de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, utilizando técnicas de MD.

Para llevar a cabo el estudio de minería de datos se seleccionó algunas preguntas de la encuesta que abarcan algunos temas sobre la profesionalización de la labor docente del personal académico, en temas de metodologías educativas, pedagógicas, psicopedagógicas, técnicas didácticas, entre otras, dando como resultado evaluaciones que muestran deficiencias en estos temas. Captar, preparar y retener al personal docente con vocación, conocimientos y es una de las tareas que la Facultad de Ingeniería (FI) debe llevar a cabo en su proceso de selección y promoción del profesorado [20].

### **1.2.2 Objetivos del proyecto**

La información, fue utilizada para realizar estudios sobre los datos, de un mismo profesor, a lo largo del tiempo y de esta forma, obtener reglas o patrones sobre su comportamiento y desempeño. Estas reglas o patrones tienen la finalidad de brindar apoyo a las divisiones y departamentos que conforman el cuerpo académico de la Facultad de Ingeniería, con el objetivo de mejorar el desempeño de los profesores que los conforman, así como obtener una mejor calidad en la enseñanza ante el grupo.

### **1.2.3 Técnicas y Herramientas**

En esta investigación se realiza un análisis de Minería de Datos sobre las encuestas que realizan los alumnos de la Facultad de Ingeniería sobre sus profesores al final de semestre, y se presenta un análisis que va de lo general a lo particular (Ver Figura 3 [20]), es decir, primeramente se analizan las calificaciones de los profesores para toda la facultad, se muestran sus tendencias y después se muestra el análisis por división y por

departamento; para ello se utilizó el software especializado WEKA y RapidMiner, así como SPSS para el análisis previo de los datos.

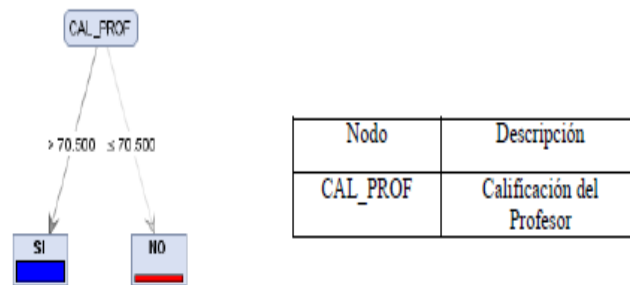


Figura 3: Fragmento modelo árbol de decisión

### 1.2.4 Perspectivas del Proyecto

La encuesta para profesores que se lleva a cabo a fin de semestre en la Facultad de Ingeniería es un instrumento que, debido al tiempo lleva aplicándose, constituye una forma de saber el comportamiento que tiene un profesor frente al grupo, y si esta información se agrupa se puede conocer el comportamiento de una asignatura, departamento o división a lo largo del tiempo, con el fin de saber qué es lo que más se le dificulta al alumno y qué tendría que hacer un profesor para mejorar el proceso enseñanza-aprendizaje. Al mejorar este proceso, se puede obtener un mejor desempeño tanto de alumnos como de profesores, y así, lograr clases más eficientes, en las que el tiempo correspondiente es totalmente aprovechado ya que el profesor conoce con anticipación, cuáles son los problemas que afectan su canal de comunicación con el alumno.

La Minería de Datos es una herramienta potente para el análisis de grandes volúmenes de datos, y es a través de ella que se pudo obtener información novedosa sobre los datos ya conocidos. Cuando esta herramienta es aplicada a las encuestas que se realizan a fin de semestre en la Facultad de Ingeniería, se obtiene información importante sobre cuáles son las preguntas de la encuesta que afectan directamente la calificación del profesor o el desempeño del alumno, es decir, si el alumno aprobó, reprobó durante su curso; con lo cual, se puede dar la interpretación de que preguntas relacionadas con la percepción del

alumno sobre su profesor o sobre su clase, contribuyen de forma significativa a su buen aprovechamiento del curso o a su escaso aprendizaje de la asignatura.

### **1.2.5 Resultados**

La aplicación de la Minería de Datos en los procesos universitarios, ayuda a enfrentar los desafíos en la mejora u obtención de una educación de calidad. La Facultad de Ingeniería (FI) de la Universidad Nacional Autónoma de México (UNAM), enfrenta situaciones como el bajo aprovechamiento escolar, la deserción, el rezago estudiantil y los bajos índices de eficiencia terminal. Los tres primeros problemas condicionan al último y todos ellos en su conjunto generan como resultado un bajo aprovechamiento de los recursos tanto humanos como económicos con los que cuenta la FI.

La educación superior requiere cumplir sus funciones con calidad y eficiencia, esto es, requiere generar los espacios adecuados para llevar a cabo una relación pedagógica rica y productiva entre el profesor y el estudiante, estableciendo un tipo de vínculo que canalice lo mejor de sus energías, capacidades e intenciones, en pocas palabras, mejorar el proceso enseñanza-aprendizaje. Para alcanzar la calidad educativa se deben aprovechar adecuadamente los recursos humanos y económicos con que cuentan las instituciones educativas.

Conjuntamente y conforme se mejore el proceso enseñanza-aprendizaje se propicia en el alumno su independencia, el logro de los objetivos educativos propuestos, favorece la adquisición de habilidades de estudio y trabajo autónomo. Como consecuencia lógica de una mejora en los métodos de trabajo del profesor, es posible esperar también una reducción en los índices de deserción y rezago, así como un incremento en el aprovechamiento escolar y una mejora en los índices de titulación [20].

### **1.2.6 Recursos utilizados**

En el presente caso de éxito se utilizaron las siguientes técnicas, algoritmos y herramientas (Ver Tabla III [20])

TABLA III: TÉCNICAS, ALGORITMOS Y HERRAMIENTAS

Técnicas	Algoritmos	Herramientas
Clasificación	Árboles de decisión, Redes Neuronales	SPSS, Weka, Rapidminer
Generación de Reglas	Reglas asociación	SPSS, Weka, Rapidminer

### 1.3 CASO DE ÉXITO 3: Obtención de Patrones y Reglas en el Proceso Académico de La Universidad de Las Ciencias Informáticas Utilizando Técnicas de Minería de Datos.

En el presente caso de éxito se aplican técnicas de minería de datos en la obtención de patrones y reglas en un proceso académico, con el fin de apoyar a la toma de decisiones institucionales.

#### 1.3.1 Introducción

En este estudio se clasifica a los estudiantes de acuerdo a su rendimiento académico, para posteriormente encontrar patrones ocultos y reglas que los caractericen; basado en las relaciones que se establecen entre el centro de procedencia de los estudiantes, nivel de escolaridad de los padres y provincia de origen con sus resultados académicos en el primer curso en la universidad [14].

#### 1.3.2 Técnicas y Herramientas

Las técnicas utilizadas dentro de este estudio son el clustering, los árboles de decisión, regresiones logísticas y algoritmos de aprendizaje inductivo, con el estudio realizado y los resultados alcanzados se pueden mejorar el proceso de formación académica y elevar la calidad de la educación en la Universidad de las Ciencias Informáticas (UCI).

#### 1.3.3 Compresión del Negocio

La universidad cuenta con una matrícula de alrededor de 10 000 estudiantes procedentes de todas las provincias y municipios del país, con los más diversos orígenes sociales y académicos; sin que, hasta el momento, se hayan realizado estudios que evalúen la influencia de estos factores en su formación posterior. Estos factores son importantes y al ser tomados en cuenta, se podría realizar el proceso de captación de los estudiantes de nuevo ingreso a la universidad, o brindarles a los ya matriculados el seguimiento

necesario, sin esto puede ocurrir en condiciones extremas a causar baja del centro. Mientras que en otros casos se dejan de identificar a los alumnos con mayor potencial, que pudieran formar parte de proyectos o grupos de investigación, o simplemente brindar a los profesores la información conveniente para que puedan brindarle atención diferenciada a sus estudiantes para fomentar el pleno desarrollo de sus capacidades y dándole así cumplimiento al objetivo primordial de la Universidad, que es el de formar profesionales cada vez mejor preparados.

Toda la información personal y docente de los estudiantes, se encuentra digitalizada y no brindan mayor utilidad que la de los reportes tradicionales.

El estudio dentro de la universidad se centra en el análisis para suplir la necesidad de contar con métodos eficientes y automáticos para explorar las grandes Bases de Datos, procesando de forma rápida y fiable la información para encontrar patrones de conocimiento apropiados para resolver un problema [14].

#### **1.3.4 Resultados**

El objetivo fundamental que se realizó en este estudio fue determinar el vínculo que existe entre el origen y procedencia social de los estudiantes de la UCI con sus resultados académicos mediante la aplicación de técnicas de agrupación y reglas de asociación de Minería de Datos. Ya en el estudio en si se orientó a la predicción de las notas de las asignaturas del primer año de los estudiantes de la UCI basado en las relaciones que se establecen entre el nivel académico de los padres, tipo de centro de procedencia, provincia y resultados académicos.

Los datos seleccionados para realizar el proyecto de Minería de Datos corresponden a la información personal y calificaciones en las asignaturas del primer año académico de los estudiantes de la UCI, y que cursan el segundo, tercero, cuarto y quinto año. Se utiliza la información del primer año académico ya que los estudiantes de primer año reciben mayor influencia de las variables a analizar como entradas para las predicciones.

El estudio además utilizó una muestra aleatoria representativa del 90% de los datos para realizar el proyecto de KDD y siguiendo los parámetros de la metodología CRISP-DM [14].



### 1.3.5 Metodología

En este estudio se desarrolló un proyecto de minería de datos guiado por la metodología CRISP-DM, para determinar la relación que existe entre la procedencia del origen social y los resultados académicos en los estudiantes de la UCI. Se construyeron, entrenaron y evaluaron los modelos de Clustering o agrupamiento y de Árboles de Decisión para obtener las reglas y patrones ocultos en los datos. Se obtuvieron modelos de predicción precisos que logran reglas con alto valor de certeza y que permiten caracterizar los datos analizados y diseños de prueba eficientes para proceder con posteriores análisis [14].

### 1.3.6 Recursos utilizados

En el presente caso de éxito se utilizaron las siguientes técnicas, metodologías y herramientas (Ver Tabla IV [14])

TABLA IV: TÉCNICAS, HERRAMIENTAS

<b>Técnicas</b>	<b>Herramientas</b>
Agrupamiento- Clúster, Árboles de decisión, Regresión, Aprendizaje inductivo.	SQLServer, Google Refine, Rapid miner, CRISP-DM

En base a la recopilación y revisión de los casos de éxito descritos se pudo tener una mejor perspectiva de los problemas solucionados en base las técnicas y herramientas seleccionados Ver Sección Resultados

## **2. CAPÍTULO II: RECOPIACIÓN DE TÉCNICAS DE MINERÍA DE DATOS.**

La Minería de Datos es un término recientemente popular, pero cabe destacar que la misma es una etapa de todo un proceso llamado proceso de extracción de conocimiento que se compone de varias fases e incorpora diferentes técnicas de los campos, como los sistemas de toma de decisión, la estadística, las bases de datos, la inteligencia artificial y otras áreas de la computación y de la gestión de la información; dentro de la Minería de Datos existen dos secciones de técnicas, las cuales son descriptivas o no supervisadas y predictivas o supervisadas. Las técnicas descriptivas permiten la identificación de comportamientos, tendencias o descubrir características significativas en los datos, mientras que las técnicas predictivas exponen o resumen las variables de interés o estiman valores futuros [20].

En este grupo se encuentran, por una parte, algoritmos que resuelven problemas de clasificación debido a que trabajan con etiquetas discretas (árboles de decisión, tablas de decisión, inducción neuronal, etc.), por otra parte, algoritmos que se utilizan en la predicción de valores como son la regresión.

El aprendizaje supervisado requiere de una fase de entrenamiento en la cual se construye un modelo de predicción y también requiere de una fase de prueba en la que se prueba la validez del modelo probándolo con todos los datos [20].

Dentro de las técnicas predictivas se encuentran:

- La Clasificación y
- La Regresión.

En las técnicas descriptivas se encuentran:

- Agrupamiento,
- Asociación

Así mismo, las técnicas de Minería de Datos para solucionar un determinado problema poseen diversos algoritmos o técnicas (Ver Tabla V [20]).

Esto significa que cada técnica tiene sus propios requisitos, y que el tipo de información obtenida con una técnica puede diferir de la obtenida con otra.

TABLA V: TÉCNICAS DE MINERÍA DE DATOS

Técnicas		Algoritmos de Minería de Datos		
<b><i>Predictivas o Supervisadas</i></b>	Clasificación	Árboles de Decisión	Redes Bayesianas	Redes Neuronales
	Regresión	Regresión logística	Algoritmos Genéticos	Generación de Reglas
<b><i>Descriptivas o No supervisadas</i></b>	Agrupación	Mapas de Kohonen.	K medias	
	Asociación	Reglas de Asociación		

En el proceso de Minería de Datos se realiza un análisis de las técnicas descriptivas o predictivas para la Extracción de Modelos o Patrones con el fin de detectar cuáles son las más eficientes y seguras para un asunto específico.

A continuación se presenta de manera resumida una descripción de cada una de las técnicas de Minería de Datos.

## 2.1 Técnicas Predictivas o Supervisadas

Son aquellas técnicas que poseen un atributo label a predecir o describir su comportamiento, conocidos previamente otros atributos llamados descriptivos. A partir de datos con etiqueta conocida, se induce un modelo que relaciona dicha etiqueta con los atributos descriptivos [20].

### 2.1.1 Árboles de decisión

Es un modelo de predicción utilizado en el ámbito de la inteligencia artificial (Ver Figura 4 [22]). Dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para

representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema [8,9], [23].

Un árbol de clasificación posee los siguientes elementos:

- Nodos intermedios: engendran dos o más (dependiendo del método empleado) segmentos descendientes inmediatos. También denominados segmentos intermedios.
- Nodos terminales: es un nodo que no se puede dividir más, también denominado segmento terminal.
- Rama de un nodo t: consta de todos los segmentos descendientes del nodo t, excluyendo t.
- Árbol de decisión completo (*T max*): árbol en el cual cada nodo terminal no se puede ramificar.
- Subárbol: se obtiene de la poda de una o más ramas del árbol completo *T max*.

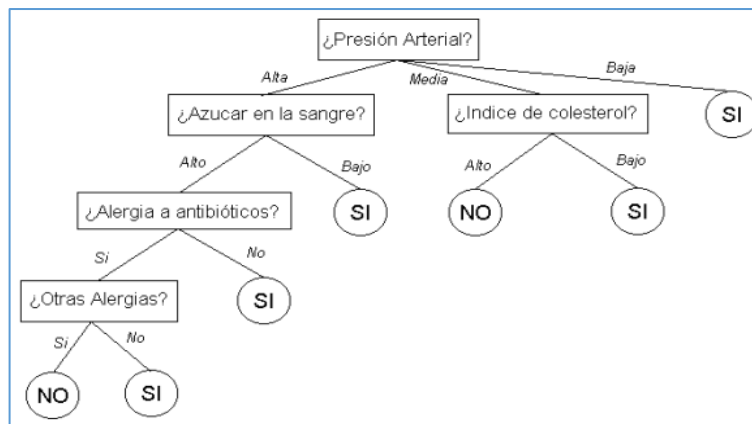


Figura 4: Técnica árbol de decisión

El ejemplo dado en la figura anterior describe si se puede llegar a recetar un fármaco x en una persona con presión arterial. Aunque esta técnica es de mayor utilidad para situaciones en que el riesgo está presente también es empleada en condiciones de certeza.

Los árboles de decisión proveen una perspectiva gráfica de la toma de decisiones, especificando las variables que son evaluadas, las acciones que deben ser tomadas y el orden en que la toma de decisión se realizará. Cada vez que se ejecuta un árbol de

decisión, sólo un camino será seguido, dependiendo del valor actual de la variable evaluada. Esto se debe a que representa, en forma secuencial, las condiciones y acciones.

Los algoritmos más utilizados al momento de generar un árbol dentro de la minería son:

- **ID3**

Es un algoritmo que prefiere árboles sencillos frente a árboles más complejos ya que, en principio, aquellos que tienen sus caminos más cortos denotan hojas más útiles a la hora de clasificar. En cada momento se ramifica por el atributo de menor peso y el proceso se repite recursivamente sobre los subconjuntos de casos de entrenamiento correspondientes a cada valor del atributo por el que se ha ramificado [23].

El ID3 construye un árbol de decisión mínimo, para un conjunto de objetos y que la información se conserve en forma organizada y entendible para cualquier persona, hace uso de una secuencia de preguntas, donde cada una de las preguntas es evaluada con el propósito de obtener la mejor respuesta posible [22-24], [26].

Este modelo es un algoritmo de aprendizaje inductivo, que busca establecer leyes o principios sobre la base de la observación de varios o todos los componentes de un conjunto o clase. Aplica la estrategia “divide y vencerás” para hacer la clasificación de los objetos y asociarlos a una clase, tomando en cuenta los valores de los atributos de los objetos. Un árbol de decisión toma como entradas objetos o situaciones que se caracterizan mediante un conjunto de propiedades [22-24], [26].

- **C4.5**

Este algoritmo genera de árboles de decisión con datos tanto nominal como datos numéricos. Permite usar como regla de división la ganancia de información, el índice de diversidad de Gini o el criterio de proporción de ganancia. Además, incluye la posibilidad de realizar una post-poda pesimista del árbol [23], [26].

Este algoritmo es una modificación del algoritmo ID3, es decir presenta más ventajas con respecto al anterior ya que permite construir arboles con valores continuos.

- **REPTree**

Este algoritmo tiene un funcionamiento en dos fases (datos de aprendizaje y datos de poda). Primero se crea un conjunto de reglas que se sobre ajusta a los datos usados para el aprendizaje, después se poda el conjunto de reglas usando ejemplares que no participaron en el aprendizaje [24].

- **CHAID**

El algoritmo CHAID (Chi Squared Automatic Interaction Detection) forma parte de los modelos llamados árboles de decisión y se aplica en la clasificación y predicción. El algoritmo parte del nodo raíz y se divide en nodos descendientes hasta llegar a los nodos hoja, donde finaliza la ramificación.

La división termina cuando el mejor valor ya no se encuentra bajo el umbral determinado y los nodos hoja del árbol son aquellos que no han sufrido ramificaciones, CHAID permite una construcción sencilla del modelo, que permite manejar variables independientes tanto discretas como continuas [25].

- **Aspectos a considerar**

Se recomienda el uso del árbol de decisión cuando el número de acciones es pequeño. Básicamente, consiste en asignar probabilidades a eventos en condiciones de riesgo o incertidumbre mediante la representación gráfica que ilustra cada estrategia o alternativa a través de una ramificación, parecidas a las ramas de un árbol. Los vértices o nodos representan los eventos de decisión, utilizando un cuadro. Los efectos derivados de la decisión se denominan acontecimientos y se representan por medio de un círculo [22].

### **2.1.1.2 Estructura del árbol de decisiones**

La construcción del árbol de decisión empieza escribiendo en la parte superior, a modo de título, la decisión o el problema que necesitará la toma de decisiones. Desde el lado izquierdo hacia el derecho se dibujarán líneas que parten desde un mismo origen que representarán las posibles decisiones a tomar.

Para su construcción y estructura es necesario considerar las distintas alternativas o cursos de acción y los posibles eventos asociados a cada curso de acción. Dentro de un árbol de decisión un nodo se toma como punto de partida, es decir, el punto desde el cual se fija un curso de acción; y se va desglosando hacia los posibles eventos asociados al curso de esa acción (Ver Figura 5 [25]).

Siguiendo el siguiente gráfico comprenderemos la estructura del Árbol de Decisiones.

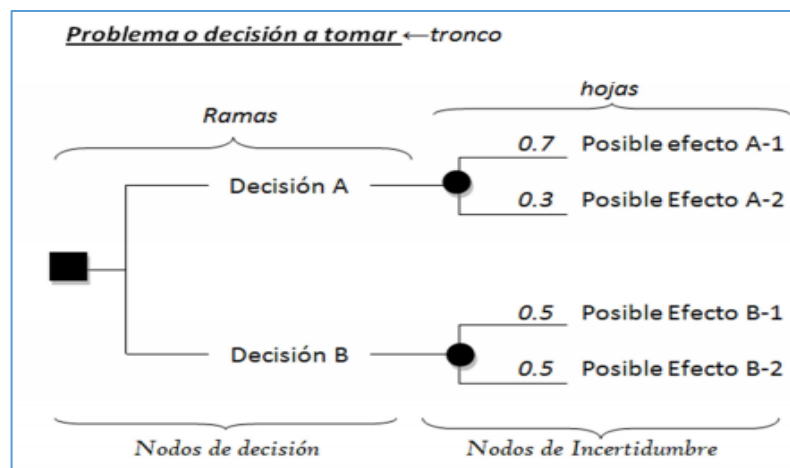


Figura 5: Estructura del árbol de decisión.

Como podemos observar en el gráfico, existen partes dentro de la estructura del árbol de decisiones, la diferenciación de estas partes ayudará a no tener confusiones al momento de evaluar los resultados del análisis del árbol de decisiones, estas partes se clasifican de dos formas:

La primera llama Tronco o nodo inicial al que lleva a las toma de decisiones; las ramas son las diferentes posibilidades de decisiones a tomar, estas ramas pueden tener más de un niveles y ser llamadas sub-ramas o ramas de primer, segundo, tercer, o "n" orden de acuerdo al nivel en que se encuentren y llamará hojas a los diferentes posibles efectos de cada decisión [25,26].

### 2.1.2 Redes Bayesianas

Las redes Bayesianas son un tipo de representación del conocimiento con incertidumbre. Este tipo de representación del conocimiento surgió de la aportación de diversas áreas de

investigación: teoría de decisión, estadística e inteligencia artificial. Representan el conocimiento cualitativo del modelo mediante un grafo dirigido a cíclico. Este conocimiento se articula mediante las relaciones de independencia/dependencia de los atributos o variables que componen el modelo [20], [28]. En la Minería de Datos se puede utilizar una representación gráfica para la especificación del modelo [20].

Las redes bayesianas permiten la representación de dependencias para razonamiento probabilístico, en la cual los nodos representan relaciones de dependencia directa entre las variables. A continuación se muestra un ejemplo (Ver Figura 6 [27]). Dada de una red bayesiana (RB) que involucra la medicina.

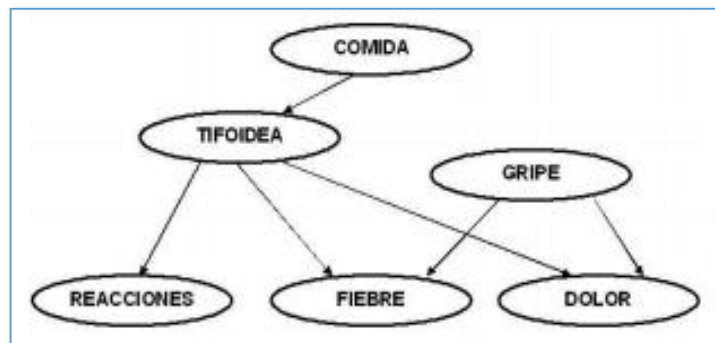


Figura 6: Ejemplo de una red bayesiana

En este caso los nodos representan enfermedades, síntomas y factores que causan algunas enfermedades. La variable a la que apunta un arco es dependiente de la que está en el origen de este, por ejemplo fiebre depende de tifoidea y gripe. La topología o estructura de la red da información de las dependencias probabilísticas entre las variables. La red también representa las independencias condicionales de una variable o conjunto de variables dadas otras variables. Por ejemplo, en la red de la figura anterior, reacciones es cond. indep. De C, G, F, D dado tifoidea. Dónde: C es comida, T es tifoidea, G es gripe, R es reacciones, F es fiebre y D es dolor. Esto es:

$$P(R|C,T,G,F,D) = P(R|T)$$

Esto se representa gráficamente por el nodo T separando al nodo R del resto de variables.



### **2.1.2.1 Naive Bayes**

Las redes bayesianas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento basados en las teorías probabilísticas (teorema de Bayes) capaces de predecir el valor de variables no observadas y explicar las observadas.

Naive Bayes es el modelo más simple de clasificación con redes bayesianas, ya que asume independencia entre todos los atributos dada una clase. NB es, por tanto, un modelo de atributos independientes. En este caso, la estructura de la red es fija y sólo es necesario aprender los parámetros (distribución de probabilidades).

El fundamento principal del clasificador NB es la suposición de que todos los atributos son independientes conocido el valor de la variable clase.

A pesar de que asumir esta suposición en el clasificador NB es sin duda bastante fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados. Además sus resultados son altamente competitivos contra los resultados de otras técnicas [29].

### **2.1.3 Redes neuronales**

Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Genéricamente, son métodos de proceso numérico en paralelo, en el que las variables interactúan mediante transformaciones lineales o no lineales, hasta obtener unas salidas. Estas salidas se contrastan con los que tenían que haber salido, basándose en unos datos de prueba, dando lugar a un proceso de retroalimentación mediante el cual la red se reconfigura, hasta obtener un modelo adecuado [9], Esta técnica es útil para realizar predicciones (Ver Figura 7 [30]).

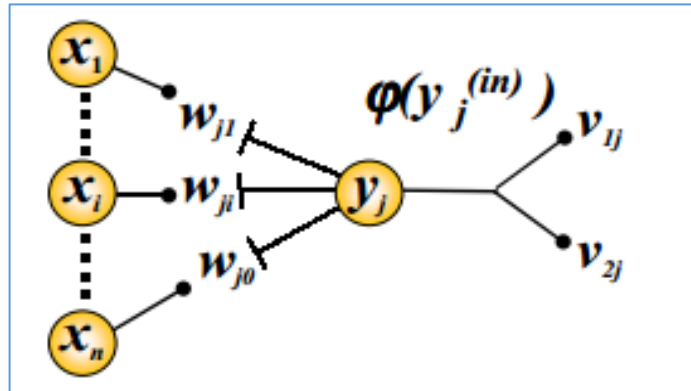


Figura 7: Técnica red neuronal.

En la figura se muestra un esquema de una red neuronal sencilla, la neurona de interés es  $y_j$ . Las  $n$  neuronas  $x_i$  están enviando señales de entradas, que son valores numéricos. Los valores  $w_{ji}$  representan los pesos sinápticos en las entradas de  $y_j$ . La notación refiere: el primer índice denota a la neurona hacia donde se dirige la información, el segundo índice denota de que neurona procede la información.

Lo que hace cada peso sináptico es simplemente multiplicar a su entrada correspondiente y define la importancia relativa de cada entrada. Cabe destacar que en una neurona se suman las entradas provenientes de todas las dendritas. Entonces se tiene que la entrada total a la neurona  $y_j$  es:

$$y_j^{(in)} = \sum_{i=1}^n w_{ji} x_i$$

En donde el índice (in) denota input o entrada. Lo que hace esto es aplicar una función de activación sobre  $y_j^{(in)}$ .

#### 2.1.4 Regresión Logística

Modelo de regresión que explica el comportamiento de una variable dependiente discreta,  $Y$ , dicotómica en función de una o más variables independientes cualitativas o cuantitativas.

Los valores que toma la variable dependiente son:

$$Y=1 \text{ e } Y=0$$

El objetivo es construir un modelo capaz de describir el efecto de los cambios de las variables explicativas sobre la probabilidad de que  $Y$  valga 1 (probabilidad del suceso de interés que denominamos éxito).

Sea  $p=P (Y=1)$  la probabilidad de que ocurra el suceso de interés. Por ejemplo, probabilidad de que un trabajador sufra un accidente;  $1-p=P (Y=0)$  es la probabilidad del suceso contrario, que denominamos fracaso (no sufra accidente) [31].

Es decir dada una variable independiente  $X$ , el modelo de regresión logística simple es:

$$\text{logit} = \ln \frac{p}{1-p} = \beta_0 + \beta_1 x$$

#### **2.1.4.1 Ejemplo de Suscripción de un diario.**

Un diario local quiere saber cuál será la probabilidad de que las personas de la ciudad se suscriban o no al servicio brindado por la empresa. Para esto se toma una muestra aleatoria de 1000 consumidores en donde se toma en cuenta la edad y sexo de los mismos (Ver Tabla VI).

Lo primero que se debe realizar es identificar tanto las variables dependientes como independientes:

**Variables:**

Subscribe: *variable dependiente* 0 no, 1 si

Edad: *variable independiente*

Sexo: *variable independiente* 0 femenino, 1 masculino

**Notación**

$$p = B_0 + B_1 * X + B_2 * X$$

$$\text{Subscribe} = B_0 + B_1 * \text{edad} + B_2 * \text{mujer}$$

TABLA VI: DATOS EMPLEADOS EN REGRESIÓN LINEAL

consumidor	Sexo	edad	Subscribe
1	0	27	1
2	0	15	0
3	1	40	1
4	1	35	1
5	1	30	1
.....+n	.....+n	.....+n	.....+n

Se corre una regresión logística con ayuda del programa gretl dando el siguiente resultado (Ver Tabla VII)

TABLA VII: RESULTADOS DE REGRESIÓN LOGÍSTICA

coefficient	
Const	-26.4653
Edad	0.787213
Mujer	-0.557795

La tabla describe que a mayor edad mayor probabilidad de subscripción, y al ser mujer la probabilidad de subscripción es menor a diferencia de los hombres. Se puede concluir que los consumidores que más se subscriben son hombres de edad mayor.

### 2.1.5 Algoritmos genéticos

Son métodos numéricos de optimización, en los que aquella variable o variables que se pretenden optimizar junto con las variables de estudio constituyen un segmento de información (Ver Figura 8 [24]). Aquellas configuraciones de las variables de análisis que obtengan mejores valores para la variable de respuesta, corresponderán a segmentos con mayor capacidad reproductiva. A través de la reproducción, los mejores segmentos perduran y su proporción crece de generación en generación. Se puede además introducir elementos aleatorios para la modificación de las variables (mutaciones). Al cabo de cierto número de iteraciones, la población estará constituida por buenas soluciones al problema

de optimización, pues las malas soluciones han ido descartándose, iteración tras iteración [9].

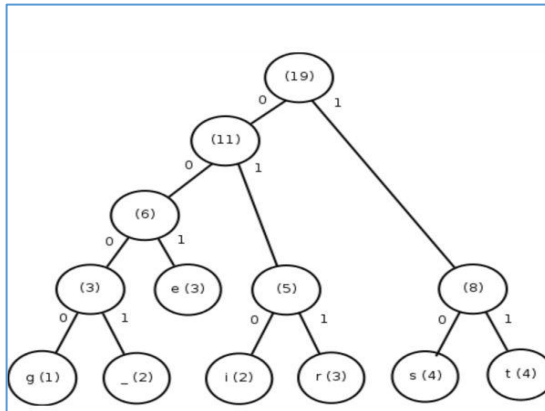


Figura 8: Técnica algoritmos genéticos.

### 2.1.6 Generación de Reglas

La generación de reglas en la minería de datos ha sido abordada en muchos estudios entre los cuales se tiene el algoritmo JRip, PART, NNge el cual permite la inducción de reglas a partir de un conjunto de datos.

A la vez las reglas generadas permiten la interpretación de conocimiento asociado a las reglas del negocio u objetivos propuestos.

- **Inducción de Reglas:**

Las reglas que genera este algoritmo tienen la ventaja que son fáciles de entender, funciona de forma similar a la regla de aprendizaje proposicional llamado algoritmo de poda incremental de error reducido (RIPPER Repeated Incremental Pruning to Produce Error Reduction), la diferencia con el algoritmo JRip es que genera reglas basándose únicamente en la ganancia de información

En la fase de crecimiento, para cada regla se añaden con glotonería condiciones a la regla hasta que la regla es perfecta (es decir, 100% de precisión). El procedimiento trata todos los valores posibles de cada atributo y selecciona el estado con la mayor ganancia de información.

En la fase de ciruela pasa, para cada regla de cualquier secuencia de finales de los antecedentes se podan con la poda métrica  $p / (p + n)$  [32].

- **JRip**

Es un algoritmo basado en el algoritmo de poda incremental de error reducido, usado en diferentes estudios de minería de datos, implementa un generador de reglas proposicionales con una nueva métrica que guía la fase de poda, una nueva condición de parada y con técnicas que optimizan las reglas aprendidas. Aunque las reglas son muy comprensibles existe la desventaja de pocas reglas generadas [32]

- **PART**

El presente algoritmo utiliza los mismos mecanismos que el C4.5 para la construcción del árbol de decisión pero evita el paso de optimización de sus reglas ya que genera una lista de decisión sin restricciones basándose en el procedimiento divide y vencerás lo que produce que éste algoritmo sea más rápido.

Es decir divide los datos generando reglas, para elaborar una regla se construye un árbol de decisión podado para el conjunto de instancias en cuestión, se toma la hoja que cubra el mayor número de instancias y se transforma en regla, y se descarta el árbol [33].

- **NNge**

Es un algoritmo basado en el método del vecino más cercano, e inducción de reglas este método utiliza la heurística para realizar búsquedas y encontrar relaciones entre los datos, para con esto generar reglas de decisión del tipo “Si, entonces” Para determinar el vecino más cercano se aplica la función de distancia Euclidiana.

Aprende de forma incremental, el algoritmo clasifica en primera instancia y luego generalizando cada nuevo ejemplo. La generalización consiste en agrupar la nueva instancia con el ejemplar de la misma clase más próxima. [33].

## 2.2 Técnicas Descriptivas o No Supervisadas

Son técnicas que tienen permiten la identificación de comportamientos, tendencias o descubrir características significativas en los datos

### 2.2.1 Agrupamiento o Clustering

Divide la base de datos en diferentes grupos. La meta del clustering es encontrar grupos que son diferentes entre sí y cuales miembros son similares el uno del otro (Ver Figura 9 [22]). Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia. Permite la clasificación de una población de individuos caracterizados por múltiples atributos (binarios, cualitativos o cuantitativos) en un número determinado de grupos, con base en las semejanzas o diferencias de los individuos. Se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes [9].

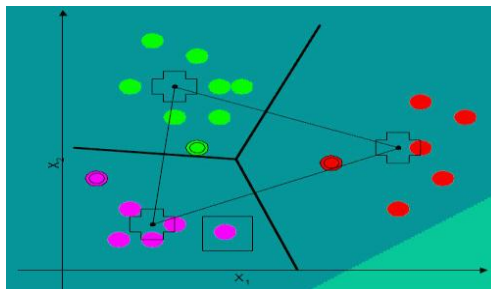


Figura 9: Técnica Clustering.

La agrupación o Clustering posee varios métodos, pero se pueden destacar los siguientes:

- ✓ **Mapas auto-organizativos de Kohonen.** Este método fue desarrollado en un principio como una red neuronal de dos capas. El método consta de una capa de entrada en donde se introducen los ejemplos, y una de competición en la que cada célula representa un prototipo.
- ✓ **K medias.** Se trata de un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar, sin etiquetar. Es el método más popular de los métodos de

agrupamiento denominados “por partición”, en contraposición de los métodos jerárquicos, los cuales parten de tantos grupos como individuos haya y van agrupando hasta que todos los elementos se encuentran agrupados en un mismo conjunto [20].

### 2.2.2 Reglas de Asociación

Una regla de asociación es una expresión de la forma  $X \rightarrow Z$  donde  $X$  y  $Z$  son conjuntos de elementos. El significado: las transacciones de la base de datos que contienen  $X$  tienden a contener  $Z$ .

En minería de datos con reglas de asociación en BD se evalúa las reglas de acuerdo al soporte y la confianza de las mismas.

En reglas de asociación, la cobertura se llama soporte (*support*) y la precisión se llama confianza (*confidence*). Se pueden leer como:

$$\begin{aligned} \text{soporte}(X \rightarrow Z) &= P(X \cup Z) \\ \text{confianza}(X \rightarrow Z) &= P(Z|X) = \frac{\text{soporte}(X \cup Z)}{\text{soporte}(X)} \end{aligned}$$

Entonces las reglas de más interés serán las de mayor soporte o confianza [34]

Además las reglas de asociación en la minería de datos se utilizan para encontrar hechos que ocurren en común dentro de un conjunto de datos. Dicho de otra manera que debe ocurrir ciertas condiciones para que se produzca cierta condición.

Para encontrar estas reglas de debe considerar cada posible combinación de condiciones para que haya una consecuencia. Al hablar de reglas de asociación también cumple un rol importante la cobertura o también llamado soporte que es el número de instancias predichas correctamente y la precisión o confianza que es la proporción de numero de instancias que es aplicada a la regla [34].

$$\begin{aligned} \{Confianza\} &\rightarrow \{Soporte\} \\ \{X\} &\rightarrow \{Y\} \end{aligned}$$



### 2.2.2.1 Ejemplo de Reglas de Asociación.

En la siguiente tabla VIII [34], se propone un ejemplo de reglas de asociación para estimar qué productos compró un cliente.

TABLA VIII: TRANSACCIONES PARA PRODUCIR REGLAS DE ASOCIACIÓN

Transacciones	
No.	Productos comprados
1	Leche, Pan
2	Pan, Mantequilla
3	Cerveza
4	Leche, Pan, Mantequilla
5	Pan
6	Leche, Pan, Mantequilla

En este contexto una regla de asociación podría ser "Si un cliente compra pan y leche, entonces también compra mantequilla", formalmente:

$$(pan \wedge leche) \Rightarrow (mantequilla)$$

El interés de una regla de asociación está dado por su soporte y su confianza, entendiéndose por soporte la frecuencia de aparición en la colección de la combinación de productos involucrados en la regla. Por ejemplo para la colección mostrada en la tabla se tiene que:

$$supp((pan \wedge leche) \Rightarrow (mantequilla)) = supp(pan \wedge leche \wedge mantequilla) = \frac{2}{6}$$

Por confianza de una regla entendemos cuánto representa el soporte de la regla, del soporte del antecedente de la regla. Por ejemplo para la colección mostrada en la tabla se tiene que:

$$conf((pan \wedge leche) \Rightarrow (mantequilla)) = \frac{supp(pan \wedge leche \wedge mantequilla)}{supp(pan \wedge leche)} = \frac{2}{3}$$

Se considera que una regla es interesante si su soporte y su confianza son mayores o iguales que ciertos umbrales de mínimo soporte y mínima confianza especificados.

Con esto la asociación, es una forma muy popular de expresar patrones de datos en una base de datos. Estos patrones pueden servir para conocer el comportamiento general del problema que genera la base de datos y así, tener más información que pueda asistir a la toma de decisiones.

Una regla de asociación es una proposición probabilística sobre la ocurrencia de ciertos estados en la base de datos; a diferencia de las reglas de clasificación, en las reglas de asociación pueden aparecer en uno o más atributos del lado derecho [34].

### **2.3 Algoritmos de Minería de Datos.**

En resumen es conveniente aclarar que todas las técnicas de minería de datos especificadas anteriormente, se pueden utilizar dentro de un proyecto, ya sea con actividades que requieran de predicción, como descripción de los datos.

Estas técnicas a su vez se derivan de otras técnicas Supervisadas y No Supervisadas que intervienen en el proceso de descubrimiento de conocimiento [35].

- **Supervisados o predictivos:** predicen el valor de un atributo de un conjunto de datos, conocidos otros atributos. A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción o descripción de datos cuya etiqueta es desconocida [20], [35].
- **No supervisados o del descubrimiento del conocimiento:** con estos algoritmos se descubren patrones y tendencias en los datos actuales. El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio de ellos [20], [35].

### **3. CAPÍTULO III: HERRAMIENTAS ÚTILES DENTRO DE LA MINERÍA DE DATOS.**

Las herramientas para la extracción de conocimiento es una parte fundamental de un proyecto de data mining (minería de datos), existen herramientas que han tenido gran aceptación debido a la facilidad de análisis que presentan, o ya sea por la diversidad de recursos que poseen.

Las herramientas útiles en el proceso de minería de datos, son algunas de uso comercial como: Clementine, Enterprise Miner o DataEngine [36], y otras de uso libre [37] como: Weka, Rapidminer y Knime

Cabe mencionar que el proceso de minería de datos involucra tan solo del 15% - 20% del esfuerzo total del proceso de Descubrimiento de Conocimiento de Bases de Datos (KDD) [8], por lo tanto se eligió una herramienta que presente características adecuadas en el proceso de extracción de conocimiento de grandes BD. Se realizó un breve análisis de algunas herramientas de uso libre como:

#### **3.1 Knime**

Es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos, desarrollado sobre la plataforma Eclipse y programado, esencialmente, en Java. Posee una interfaz gráfica muy amigable de nodos y flechas que se despliegan y combinan de manera interactiva permitiendo un fácil y rápido procesamiento de datos, para el análisis, modelado y visualización [37].

Su uso se basa en el diseño de un flujo de ejecución que plasma las distintas etapas de un proyecto de minería de datos [38]. Proporciona nodos mediante los cuales se puede realizar:

- ✓ Entrada de datos (IO>Read).
- ✓ Salida de datos (IO > Write).
- ✓ Preprocesamiento (Data Manipulation), para filtrar, discretizar, normalizar, filtrar, seleccionar variables.

- ✓ Minería de datos (Mining), para construir modelos (reglas de asociación, clasificación, clustering, etc.)
- ✓ Salida de resultados (Data Views) para mostrar resultados en pantalla (ya sea de forma textual o gráfica) [38].

#### **Características:**

- ✓ Manipulación de filas, columnas, como muestreos, transformaciones, agrupaciones [39].
- ✓ Visualización (histogramas) [39].
- ✓ Creación de modelos estadísticos y de minería de datos, como árboles de decisión, regresiones, etc. [39,40].
- ✓ Validación de modelos, como curvas ROC [39,40].
- ✓ Aplicación de dichos modelos sobre conjuntos nuevos de datos [39,40].
- ✓ El carácter abierto de la herramienta hace posible su extensión mediante la creación de nuevos nodos que implementen algoritmos a la medida del usuario [38-40].
- ✓ Importación y exportación de flujos de trabajo [39,40].
- ✓ Entrada/Salida: Recupera datos de archivos o bases de datos [39,40].

### **3.2 Weka**

Es una de las herramientas para aplicación de tareas de data mining más reconocidas, que permite proceso previo, clustering o generación de grupos de datos, clasificación, regresiones, visualización y selección de propiedades [41]. Maneja lenguaje Java, es multiplataforma y está bajo una licencia GNU lo cual ha permitido que sea una de las más utilizadas en los últimos años en el área de la Minería de datos [42,43].

#### **Características:**

- ✓ Es de facilidad uso gracias a su interfaz gráfica de usuario dividida en 4 escenarios como: Explorer, Experimenter, KnowledgeFlow, SimpleCLI [42,43].

- ✓ Desarrollado en lenguaje java y permite la ejecución en cualquier plataforma [42-44]
- ✓ Para el preprocesamiento de datos consta de una amplia colección de técnicas [44].
- ✓ Soporta ficheros en formato ARFF, CSV, Excel y conexión jdbc con BD [44].
- ✓ Disponible bajo la licencia publica general GNU [43,44].
- ✓ Escenario de experimentos para pruebas estadísticas [43,44].
- ✓ Consta de flujos de datos [43,44].

### 3.3 Rapidminer

Es una herramienta muy utilizada en el proceso de minería de datos la cual permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico [45,46]. Rapidminer cuenta con dos componentes:

**RapidMiner:** Versión stand-alone para analistas. Implementa todos los operadores de data mining, modelos predictivos, modelos descriptivos, transformación de datos, series de tiempo, etc. [46].

**RapidAnalytics:** Versión Servidor de RapidMiner. Permite trabajo colaborativo, escalable y concurrente de múltiples usuarios, capacidad de delegar en bases de datos (In-Database Mining) y otras mejoras de funcionalidad como: plataforma Web de publicación de informes, implementación de sistemas de scoring, diseño y navegación Web de informes, Single-sign on e integración vía Servicios Web, entre otras [46].

#### Características:

- ✓ Es multiplataforma [46,47].
- ✓ Representación interna de los procesos de análisis de datos en ficheros XML y jerarquización de operadores [47,48].
- ✓ Permite a los experimentos componerse de un gran número de operadores anidables arbitrariamente, que se detallan en archivos XML [47].

- ✓ Incluye gráficos y herramientas de visualización de datos [22-25] (ver anexo 3).
- ✓ Dispone de un módulo de integración con R [22-24].
- ✓ Es un software de tipo Open-Source con licencia GNU GPL, basado en java [47].

## e. Materiales y Métodos

Antes de emplear una metodología en la presente investigación se realizó un análisis comparativo de las metodologías más utilizadas, características y beneficios que poseen. Entre las metodologías con mayor aceptación y uso se tiene a CRISP-DM, SEMMA y Catalyst (conocida como P3TQ) [49]

La importancia para utilizar metodologías dentro de la Minería de Datos surge de la necesidad de contar con un proceso o pasos estándares para la resolución de problemas con las herramientas y técnicas disponibles dentro de la minería.

**SEMMA:** desarrollado por el Instituto de soluciones analíticas SAS, se caracteriza principalmente por priorizar sus fases desde un punto de vista técnico, es decir, dando prioridad a las prácticas usadas para su implementación y obtención de resultados. El nombre de esta terminología es el acrónimo correspondiente a las cinco fases básicas del proceso: Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado), Assess (Valoración).

Sample: es la primera etapa del proyecto. En ella se prepara los datos para su posterior exploración [50].

Explore: se trata de la exploración de los datos. Es una de las partes más trabajosas pero también la más interesante de las etapas establecidas.

Modify: esta etapa se centra en la selección y transformación de variables y datos que servirán para la construcción de los modelos. Entre otras tareas a realizar destacan: la reducción de dimensión, imputación de valores

Model: es la generación de modelos con la aplicación de algoritmos como regresión logística, árboles de decisión, análisis factorial discriminante, redes neuronales.

Assess: se realiza la comparación de modelos generados mediante análisis del diagrama ROC. [50]

**CATALYST:** conocida como P3TQ (Product, Place, Price, Time, Quantity), fue propuesta por Dorian Pyle en el año 2003 esta metodología se divide en dos etapas, la primera denominada Modelado de Negocio o MII, y la segunda llamada Explotación de Información o MIII.

El Modelo de Negocio (MII), proporciona una guía de pasos para identificar un problema y los requerimientos reales de la organización.

El Modelo de Explotación de Información (MIII), proporciona una guía de pasos para la construcción y ejecución de modelos de minería de datos a partir del Modelo [51]

**CRISP-DM:** fue creada por el grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000. Esta metodología constituye la guía de referencia más utilizada en el desarrollo de proyectos de Explotación de Datos. Posee una serie de fases las cuales son: Comprensión del Negocio o Problema, Compresión de los Datos, Preparación de los datos, Modelado, Evaluación y Validación, Implantación detalladas más adelante [52].

A continuación se presenta una tabla comparativa de las metodologías estudiadas (Ver Tabla IX)

TABLA IX: METODOLOGÍAS MD

	<b>SEMMA</b>	<b>CATALYST</b>	<b>CRISP-DM</b>
Permite elección libre de herramientas	NO	SI	SI
Cantidad de fases	5	6	5
Las fases pueden relacionarse	NO	SI	SI
Posee información de cada fase	NO	SI	SI
Utilizada en proyectos empresariales y académicos	Bajo	Medio	Alto

Con el análisis realizado se puede concluir que la metodología SEMMA se centra en los aspectos técnicos de los proyectos de explotación de datos. Además está acotada ya que ha sido diseñada para ser implementada con los productos SAS.

CRISP-DM es más completa y abierta que SEMMA, pero no llega al detalle de la metodología CATALYST, ya que nombra etapas del proceso de la explotación de la información, pero no se analizan los pasos, resultados y situaciones que se pueden dar dentro de cada etapa. En la siguiente figura se presenta las estadísticas de las metodologías más utilizadas (Ver Figura 10).



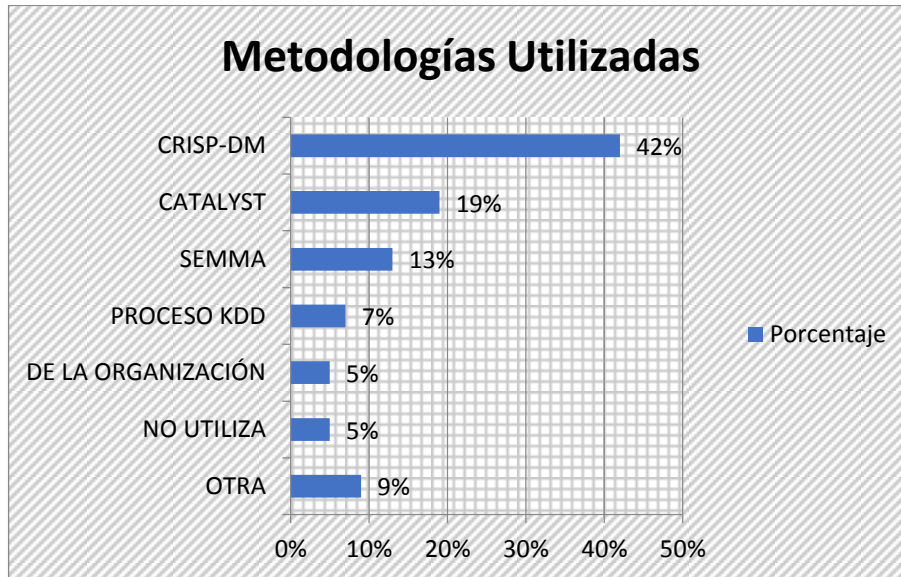


Figura 10: Encuesta realizada por la KDnuggets Metodologías utilizadas en MD.

El gráfico muestra que la técnica más utilizada es CRISP-DM seguida de CATALYST dentro de la encuesta realizada, por lo tanto y por lo especificado en el análisis de las mismas el desarrollo del presente trabajado de titulación, se basó la metodología, que ofrece en detalle los pasos y procesos que se debe seguir para la culminación de cualquier proyecto de MD, de manera satisfactoria, la cual es Cross Industry Standard Process for Data Mining (CRISP-DM) [49-52], muy utilizada en el campo académico y empresarial.

CRISP-DM presenta una flexibilidad desglosada en una sucesión de fases, es decir cada fase contiene varias tareas generales. Estas tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas (Ver Figura 11 [52])

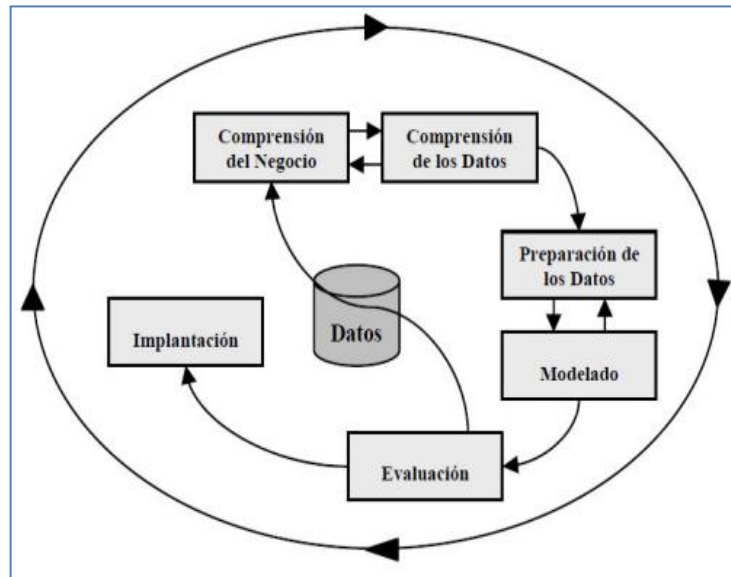


Figura 11: Modelo de proceso CRISP-DM

En cuanto al desarrollo y alcance de los objetivos propuestos se ejecutó cada etapa o fase dentro de la metodología CRISP-DM, estas fases son 6 las cuales se detallan a continuación [49-52]:

### Fase 1: Comprensión del Negocio o Problema

La etapa inicial de la investigación consistió en determinar y definir claramente los problemas a los que se va a dar solución. Para ello se accedió a diversas fuentes bibliográficas a fin de justificar los problemas abordados. Es decir para obtener el mejor provecho de Data Mining, fue necesario entender de la manera más completa el problema que se deseaba resolver (ver sección Resultados: Apartado 3.1.1), esto permitió recolectar los datos correctos e interpretar correctamente los resultados. En esta fase se realizó varias tareas como:

- **Determinar los objetivos del negocio:** esta tarea permitió al analista entender cuál es la necesidad que se deseó abarcar.
- **Evaluación de la situación:** en esta tarea se realizó un análisis más detallado para encontrar todos los recursos, restricciones y suposiciones que deben ser considerados al establecer los objetivos del análisis de datos y la meta del proyecto.

- **Determinación de metas de la minería de datos:** por medio de esta tarea se establecieron los objetivos del proyecto desde el punto de vista técnico del data mining.
- **Elaboración del Plan del proyecto:** se elaboró el plan que posee una secuencia de pasos a realizar durante la investigación, incluyendo la selección de herramientas y técnicas.

## **Fase 2: Compresión de los Datos**

La segunda fase, corresponden a la comprensión de los datos, la recolección inicial de los mismos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes

La comprensión de los datos abarca el entendimiento total de la estructura, creación de funciones, vistas y consultas que permitan a su vez el análisis de la información y registros que contiene la BD disponible.

Dentro de esta fase se llevaron las siguientes tareas:

- **Recolección Inicial de los datos:** para realizar el proceso de minería de datos, fue necesario que previamente se decida y seleccione los datos que intervendrían en el tratamiento de la información (Ver sección Resultados apartado 1).
- **Descripción de los datos:** se realizó la descripción de los datos que fueron adquiridos incluyendo su formato.
- **Exploración de los datos:** en esta etapa se realizó una exploración de los datos y en base a ello un análisis estadístico acerca del proceso de evaluación.
- **Verificación de la calidad de los datos:** se verificó la calidad de los datos con la finalidad de corregir errores, completar campos nulos o en su defecto eliminarlos, esto con el objetivo de realizar el proceso de limpieza (ver sección Resultados: Apartado 3.1.2).

### **Fase 3: Preparación de los datos**

En esta fase y una vez efectuada la recolección inicial de datos, se procedió a su preparación para adaptarlos a las técnicas de Data Mining que se utilizaron posteriormente, tales como técnicas de visualización de datos, u otras medidas para exploración de los datos. La preparación de datos incluye las tareas generales de selección de datos a los que se aplicó una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato (ver sección Resultados: Apartado 3.1.3). Las tareas que se llevaron a cabo fueron:

- **Selección de los Datos:** esta tarea constó en decidir qué datos iban a ser utilizados para el análisis. Criterios a tener en cuenta pueden ser la relevancia para el objetivo del proyecto y el límite del volumen de datos con el cual se puede trabajar.
- **Limpieza de los Datos:** El objetivo fue preparar los datos de una forma adecuada para las técnicas de análisis seleccionadas. Por otro lado también se corrigió los problemas de calidad detectados en las etapas anteriores. Esto implica por ejemplo elegir subconjuntos de datos o insertar valores.
- **Construcción de Datos:** dentro de la construcción, se armó las estructuras de datos que permitió llevar a cabo la aplicación de técnicas y obtención del modelado.
- **Integración de datos:** aquí se combinaron datos de diferentes fuentes creando nuevos registros.

### **Fase 4: Modelado**

En esta fase se seleccionaron las técnicas de modelado más apropiadas para el proyecto de Data Mining específico. Las técnicas a utilizar en esta fase se eligieron en función de los siguientes criterios: ser apropiada al problema, disponer de datos adecuados, cumplir los requisitos del problema, tiempo adecuado para obtener un modelo y conocimiento de la técnica. Esta fase involucra las siguientes tareas:

- **Selección de técnica de modelado:** el objetivo de esta tarea fue elegir las técnicas de modelado útiles sobre el set de datos.
- **Generación del diseño de pruebas:** el objetivo fue diseñar las pruebas para probar la calidad y validez de los resultados arrojados por el modelo.
- **Construcción de Modelo:** en esta sección se explicó los resultados que se consiguieron con la aplicación de los diferentes algoritmos, los pasos y procedimientos que se tomaron en cuenta para llegar a la construcción de los modelos propuestos.
- **Evaluación de modelos:** en esta tarea se hizo un análisis comparativo que permitió determinar el grado de concordancia o la confiabilidad en cada uno de los algoritmos aplicados, cabe señalar que no se evaluó el resultado de los modelos si no que se analizó el desempeño y calidad de los mismos (ver sección Resultados: Apartado 3.1.4).

#### **Fase 5: Evaluación y Validación.**

En esta fase se evaluó 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 aplicó solamente para los datos sobre los que se realizó el análisis. Esta fase también abarcó la evaluación y revisión de los resultados.

Las tareas llevadas a cabo dentro de esta fase fueron:

- **Evaluación de Resultados:** en esta etapa se evaluaron los resultados del modelo desde la perspectiva del negocio. Se analizó el nivel en el que el resultado se acercó a los objetivos y se trató de determinar si por alguna razón el modelo es deficiente.
- **Evaluación de los resultados de la minería de datos con respecto a los Factores Críticos del Éxito:** esta etapa permitió evaluar los resultados en base a los objetivos propuestos (ver sección Resultados: Apartado 4.2.1).

## **Fase 6: Implantación**

Normalmente los proyectos de Data Mining no terminan en la implantación del modelo, sino que se deben documentar y presentar los resultados de manera comprensible en orden a lograr un incremento del conocimiento. Es decir dar un informe de los resultados obtenidos y estos puedan ser útiles para la automatización de procesos o apoyo a la toma de decisiones [53].

- **Producción del Informe Final:** Este informe puede ser tanto un resumen del proyecto como también una presentación final de los resultados del data mining, en este caso es la descripción en detalle de la memoria final del presente trabajo de titulación (Ver sección Resultados: Apartado 5.1).

## **f. Resultados**

El desarrollo del trabajo de titulación estuvo inmerso en un conjunto de fases, las cuales permitieron llevar a cabo cada actividad, objetivos, procesos, alcance y propuesta en su totalidad de manera ordenada y objetiva.

A continuación se presenta en detalle cada una de ellas:

### **1. FASE 1: Recopilación y Análisis de los Datos proporcionados con respecto al proceso de Evaluación al Desempeño Docente.**

#### **1.1 Realizar entrevistas con el responsable de la Unidad de Telecomunicaciones e Información (UTI) y con la coordinadora del proceso de evaluación de desempeño docente, para solicitar autorización en cuanto a manipulación de los datos y acceso a los mismos.**

En este primer aspecto se llevó a cabo conversaciones con el responsable de la Unidad de Telecomunicaciones con respecto a la disposición de los datos para su posterior utilización. Además se realizó una exposición de la propuesta de investigación a la coordinadora del proceso de evaluación de desempeño docente y a la responsable de Estadística y Base de Datos de la Unidad de Evaluación institucional, esto con el fin de evaluar la propuesta, de tal manera que no exista riesgo alguno en cuanto a la confidencialidad e integridad de los datos existentes en dicho proceso. Adicionalmente se elaboró un acta de confidencialidad dirigida hacia la coordinadora del proceso de evaluación (Ver Anexo 1).

Ya con esto se obtuvo el consentimiento por parte de la coordinadora del proceso de evaluación, para poder acceder a los datos almacenados, para lo cual se adquirió un certificado de acceso a los datos, de Evaluación de Desempeño Docente de la Universidad Nacional de Loja (Ver Anexo 2). En el cual se señala la autorización para acceder a la información en el desarrollo del Trabajo de Titulación.

Además se obtuvo permiso para el acceso al Web Services de la Universidad con el propósito de obtener los datos en cuanto a las notas de los estudiantes en los periodos académicos de estudio (Ver Anexo 3).

## **1.2 Recopilación de los datos necesarios dentro del proceso de evaluación al desempeño docente, en el formato requerido.**

Los datos se obtuvieron directamente de la Unidad de Telecomunicaciones e Información (UTI); mediante el acceso al Web Services para la obtención de las notas de los estudiantes y en cuanto al proceso de Evaluación de Desempeño Docente (EDD) a través de un CD, donde constan datos como:

- Campos descritos por los estudiantes en cuanto a la opinión dada hacia un docente evaluado en un determinado módulo.
- Calificaciones de los docentes del AEIRRNR evaluados en cada módulo y periodo académico correspondiente desde el 2011 hasta el 2013.
- Parámetros o componentes calificados.
- Cuestionarios de evaluación.

## **1.3 Explorar y analizar los datos, para la toma de decisiones adecuadas en la generación de un Modelo de MD.**

El siguiente apartado se lo ha dividido en 2 aspectos los cuales son:

- ✓ La Exploración de la Información y
- ✓ La Disposición de los datos.

### **1.3.1 Exploración de la Información.**

En este punto se presenta la información que se debió conocer antes de proceder al procesamiento de datos, fue necesario conocer la situación actual, identificación de problemas, marco legal, en si conocer a fondo el escenario de estudio

#### **1.3.1.1 Modelo de evaluación en la Universidad Nacional de Loja.**

Los procedimientos, para identificar fortalezas, debilidades y la valoración ponderada de los resultados dentro de la evaluación de desempeño docente; está basada en técnicas utilizadas para recopilar la información como encuestas y cuestionarios.



Las personas que informan, están consideradas como claves y son los estudiantes, docentes y la Comisión Académica conformada por el directivo y el Par Académico de la Carrera o Programa de Posgrado de la Universidad Nacional de Loja (Ver Tabla X) [1].

### 1.3.1.2 Informantes de Calidad

En el análisis y valoración del desempeño docente participan los involucrados en el proceso de enseñanza aprendizaje: estudiantes, coordinadores de módulo, coordinadores de carrera y docentes en general.

TABLA X: INFORMANTES DE CALIDAD.

No.	DENOMINACIÓN
01	Estudiantes
02	Docentes
03	Comisión Académica de la Carrera o Programa de Postgrado

### 1.3.1.3 Delimitación del Objeto a Evaluar

La evaluación para medir el desempeño de un docente se basa en 3 componentes importantes, como son la Capacidad Profesional, Capacidad Pedagógica, y La Práctica de Valores (Ver Tabla XI) [1].

TABLA XI: ÁMBITO Y COMPONENTES EVALUADOS

ÁMBITO	COMPONENTES
Docencia	1. Capacidad Profesional
	2. Capacidad Pedagógica
	3. Práctica de valores

**a) Capacidad profesional.** Asume que el profesorado universitario es profesional, es decir, que tiene una formación sujeta a licencia; un amplio margen de discrecionalidad en su acción; que ésta se resuelve desde supuestos teórico-científicos; que implica un constante esfuerzo de mejora y actualización; que está orientada a cubrir necesidades sociales básicas, fundamentalmente, la docencia en Educación Superior. Es decir la actividad del profesor universitario es dar clase, lo que parece obvio, pero que debe recordarse. El profesor, sin duda, podrá investigar,

gestionar la institución, buscar recursos y otras actividades, pero su existencia es posible porque hay alumnos a los que debe enseñar y, lo que es más importante, que éstos deben aprender [1].

**b) La capacidad pedagógica.**- Ser un docente universitario competente desde una concepción humanista de la educación significa no sólo ser un conocedor de la ciencia que explica (física, matemáticas), sino también de los contenidos teóricos y metodológicos de la psicología y la pedagogía contemporáneas que lo capacite para diseñar un proceso de enseñanza-aprendizaje que potencie el desarrollo de la personalidad del estudiante. Existen aspectos importantes relacionados con la práctica pedagógica que atiendan a la dimensión ética y están vinculadas al currículo, tales como las estrategias de enseñanza-aprendizaje-evaluación, las cuales incluyen actividades, procedimientos, tareas, recursos y medios utilizados por el docente para promover aprendizajes significativos [1].

**c) La práctica de valores.** No se trata de que el docente enseñe valores, se trata de que pueda crear una atmósfera moral en el aula, dentro de la cual se use a la ética en función de los conflictos, se estimule el pensamiento de los estudiantes para que ellos construyan juicios y razones. Todo lo que ocurra dentro de un salón de clase y vaya en contra de la vida, es antiético. Se requiere entonces un docente puntual, organizado, negociador, oyente, preparado cognitivamente, con sensibilidad social, capaz de discernir, canalizar las dificultades que surjan en el recinto universitario, atendiendo al bien colectivo, más que al personal [1].

#### 1.3.1.4 Códigos para los Ámbitos, Componentes y Criterios de Análisis

Dentro del proceso de evaluación, se usan códigos de identificación para cada aspecto o componente evaluado, con el fin de facilitar los procesos de tabulación (Ver Tabla XII)

TABLA XII: CÓDIGOS PARA COMPONENTES EVALUADOS.

COMPONENTES	CÓDIGO
♦ Capacidad profesional	CPF
♦ Capacidad pedagógica	CPG

♦ Práctica de valores	PV
-----------------------	----

Todas las preguntas de los cuestionarios contienen una escala que el encuestado debe responder. (Ver Tabla XIII) [1], muestra las alternativas de las escalas de valoración cualitativa a utilizarse.

TABLA XIII: ESCALA DE VALORACIÓN CUANTITATIVA.

Totalmente de acuerdo	Mayoritariamente de acuerdo	Minoritariamente de acuerdo	En desacuerdo
4	3	2	1

Los cuestionarios son aplicados a los estudiantes, docentes y, Comisión Académica de la Carrera o Programa, que responderá en conjunto el cuestionario para valorar el desempeño de los docentes. Se evaluarán a todos los docentes que participan, sea en el módulo, curso, seminario o taller.

Los comentarios y sugerencias que se solicitan luego sirven solo para reafirmar la respuesta solicitada (Ver Anexo 4) [1].

### 1.3.1.5 Ponderación

La ponderación se realiza por componente y ponderación en base a los participantes.

- **Ponderación por Componente:** A continuación se presenta los valores ponderados de los 3 componentes, que son parte del objeto a evaluar (Ver Tabla XIV) [1]. A la capacidad profesional se le otorga el 30%, a la capacidad pedagógica el 50% y a la práctica de valores el 20%.

TABLA XIV: PESO CUANTITATIVO PARA CADA COMPONENTE.

COMPONENTE	%
1. Capacidad profesional	30
2. Capacidad pedagógica	50
3. Práctica de valores	20
<b>Total</b>	<b>100</b>

- **Ponderación de los participantes:** los participantes dentro de la evaluación son los estudiantes, la comisión académica, y docente, a continuación el peso por participantes (Ver Tabla XV) [1].

TABLA XV: PESO CUANTITATIVO POR INFORMANTES

<b>PARTICIPANTES</b>	<b>%</b>
<b>Estudiantes</b>	<b>60</b>
<b>Comisión Académica(pares Académicos)</b>	<b>25</b>
<b>Docentes</b>	<b>15</b>
<b>Total</b>	<b>100</b>

### 1.3.1.6 Valoración

La valoración de los resultados del desempeño docente, tiene un equivalente que se presenta de forma Cualitativa y Cuantitativa en los siguientes parámetros (Ver Tabla XVI) [1].

TABLA XVI: EQUIVALENTE CUALI-CUANTITATIVO DEL PROCESO EVALUATIVO.

<b>EVALUACIÓN CUALITATIVA</b>	<b>EVALUACIÓN CUANTITATIVA (%)</b>	<b>RESULTADOS</b>
A= Muy buen desempeño docente: Muy buena capacidad pedagógica, capacidad profesional, práctica de valores y compromiso institucional	81-100%	Nivel Destacado
B= Buen desempeño docente	61-80%	Nivel Satisfactorio
C= Regular desempeño docente	41-60%	Nivel Poco Satisfactorio
D= Insuficiente desempeño docente	0-40%	Nivel Insatisfactorio

### 1.3.2 Disposición de los datos.

Dentro de la disposición de los datos, se presenta un análisis de la BD proporcionada por la UTI, en la misma existen diferentes tipos de clases que ayudan en la obtención y análisis requerido por la presente investigación.

**Base de Datos Proporcionada:** La BD se encuentra dentro de un formato administrado por Postgres, y cuenta con 30 tablas, 5 vistas y la creación de funciones en el análisis y recopilación de registros pero las tablas que interesan en el proceso de análisis se describen más adelante.

Dentro de los datos proporcionados por la UTI, se encuentran 2 encuestas de satisfacción estudiantil y una de evaluación de desempeño docente, ambas encuestas son similares ya que poseen las mismas características; es decir se evalúan 3 parámetros: capacidad profesional, capacidad pedagógica y práctica de valores. La diferencia radica en que la primera posee 37 preguntas de elección y 3 de opinión y la segunda 45 de selección y 3 de opinión (Ver Anexo 4).

Cabe señalar que los datos correspondientes a periodos de evaluación anteriores se encuentran en un formato de mysql en una estructura diferente a los datos almacenados actualmente, y que los datos que se analizan dentro de la investigación están en postgres y corresponden a los periodos 2011, 2012, 2013 del AEIRNNR.

Las tablas con más relevancia dentro del proceso de análisis son las siguientes (Ver Tabla XVII, XVIII, XIX)

- **Carreras y Asignaturas**

TABLA XVII: DATOS CORRESPONDIENTES CARRERAS ASIGNATURAS.

Tabla	Descripción
app_areasga	Devuelve el nombre y número de áreas en la Universidad Nacional de Loja
app_direccioncarrera	Devuelve la carrera con su respectivo director
app_asignatura	Devuelve cada una de las asignaturas dictadas
app_asignaturadocente	Devuelve las asignaturas por docente
app_estudianteasignaturadocente	Devuelve los estudiantes que corresponden a una asignatura dictada por un docente

- **Estudiantes y Docentes**

TABLA XVIII: DATOS CORRESPONDIENTES ESTUDIANTES DOCENTES

Tabla	Descripción
app_estudianteperiodoacademico	Devuelve los estudiantes que corresponden a un determinado periodo académico

estudiante_evaluacion	Devuelve los estudiantes que corresponden a un periodo de evaluación determinado
app_docenteperiodoacademico	Devuelve los docentes que corresponden a un determinado periodo académico
docente_evaluacion	Devuelve los docentes que corresponden a un periodo de evaluación determinado

- **Cuestionario y Evaluación**

TABLA XIX: DATOS CORRESPONDIENTES CUESTIONARIO EVALUACIÓN.

<b>Tabla</b>	<b>Descripción</b>
app_cuestionario	Devuelve los cuestionarios en los periodos de evaluación
app_seccion	Devuelve las secciones que contiene un cuestionario
app_pregunta	Devuelve las preguntas dentro de una sección en un cuestionario
app_contestacion	Devuelve la contestación con respecto a una pregunta dentro de una evaluación
app_evaluacion	Devuelve el id de evaluación, con fecha de inicio y fin
app_periodeloevaluacion	Devuelve los periodos de evaluación existentes dentro de un periodo académico
app_periodeloacademico	Devuelve los periodos académicos en los que se llevan a cabo las evaluaciones.
app_ofertacademicasga	Devuelve las ofertas académicas propuestas dentro de un área.
app_tipoinformante	Devuelve estudiantes, docentes, directivo o par académico categorizado en el proceso.
app_usuario	Devuelve el tipo de usuario que ingresa al sistema

Además se recopiló los datos en cuanto a las notas de los estudiantes, existentes en web services del Sistema de Gestión Académica de la Universidad Nacional de Loja (Ver Figura 12).



Figura 12: Categorías e información dentro del Web Services.

Las notas de los estudiantes se los obtuvo de la categoría Académica (Ver Tabla XX)

TABLA XX: DESCRIPCIÓN DE LOS SERVICIOS POR CATEGORÍAS.

Categoría	Descripción
Académica	Esta categoría tiene métodos que devuelven información acerca de estudiantes y docentes.
Institucional	Esta categoría retorna información institucional, como datos de módulos, paralelos, carreras y áreas
Personal	Esta categoría retorna a través de sus métodos, información personal de estudiantes y docentes.
Validación	Esta categoría posee métodos que sirven para la validación de docentes y estudiantes dentro de la BD.
Estadística	Esta categoría se retorna información estadística, como número de estudiantes aprobados y número de estudiantes reprobados, entre otros.

En base a un análisis de los datos proporcionados, se observó que la información almacenada en cuanto al proceso de EDD, se encuentra en 2 bases de datos relacionales una en el formato de administración de mysql, cuyos periodos de evaluación corresponden a

los periodos académicos 2009-2010 y otra en Postgres que contiene datos del periodo académico 2011 hasta el 2013.

Siendo la BD en postgres la tomada para el presente trabajo de titulación, ya que cuenta con los datos suficientes y requeridos dentro de la investigación. A continuación se describen algunos detalles de la misma (Ver Tabla XXI XXII). Los datos descritos a continuación se los obtuvo mediante el gestor de base de datos postgres generando reportes y consultas (Ver Anexo 5).

TABLA XXI: PERIODOS ACADÉMICOS DE EDD.

Periodo	Nombre Periodo Académico
<b>1</b>	Septiembre 2011 - Julio 2012
<b>2</b>	Septiembre 2012 - Febrero 2013
<b>3</b>	Marzo - Julio 2013

La tabla muestra que desde el 2011 hasta la actualidad existen 3 periodos académicos en los que se llevó acabo el proceso de EDD.

TABLA XXII: PERIODOS DE EVALUACIÓN

Nro.	Periodo Académico	Periodo Evaluación
1	Septiembre 2011-Julio 2012	<b>a.</b> "Encuesta de Satisfacción Estudiantil 2012"
2	Septiembre 2012-Febrero 2013	<b>b.</b> "Evaluación Desempeño Docente 2012"
3	Marzo-Julio 2013	<b>c.</b> "Encuesta de Satisfacción Estudiantil 2013"

La tabla anterior muestra los nombres de las evaluaciones en cada periodo académico, estas evaluaciones corresponden tanto a estudiantes, docentes, directivos y par académico. Dentro del mismo análisis se puede determinar la cantidad de estudiantes por carrera que participaron en el proceso (Ver Figura 13-16).



- **Número de estudiantes por Carrera**

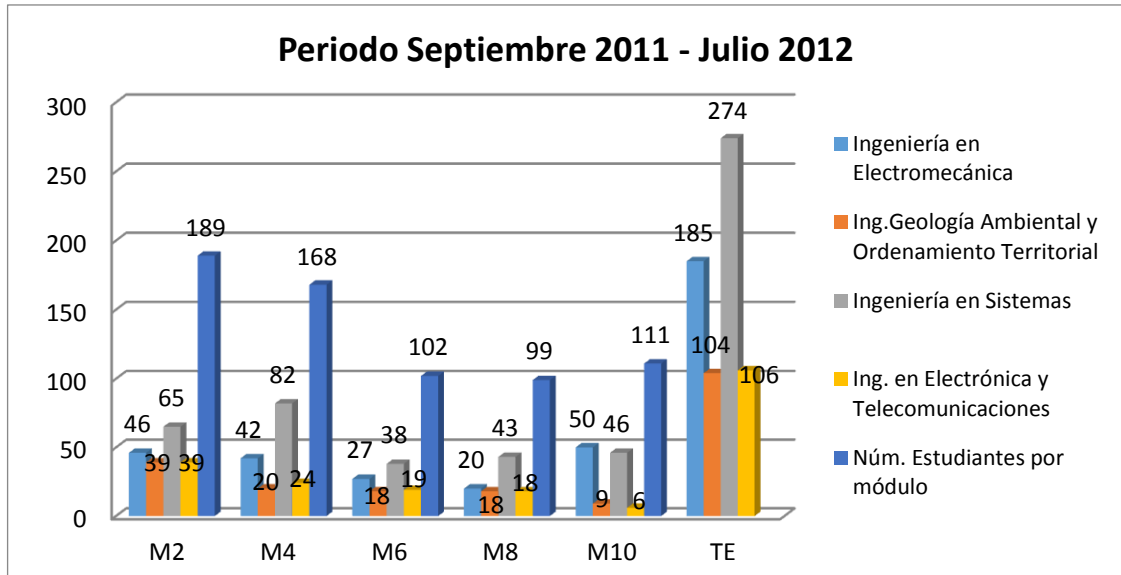


Figura 13: Estudiantes del AEIRNNR por Carrera y Módulo.

En la tabla y gráfico anterior se presenta que dentro del periodo septiembre 2011-julio 2012 la mayor cantidad de estudiantes corresponden a la carrera de ingeniería en sistemas con 274 estudiantes, seguido de la carrera de electromecánica con 185, la carrera con menos estudiantes es la carrera de Ing. Geología Ambiental y Ordenamiento Territorial con 104 estudiantes. Además el módulo que más estudiantes tiene es el módulo 2 que son 189 estudiantes, seguido del módulo 4 con 168 estudiantes. A diferencia del módulo 8 que es el módulo con menos estudiantes con tan solo 99 alumnos por periodo académico.

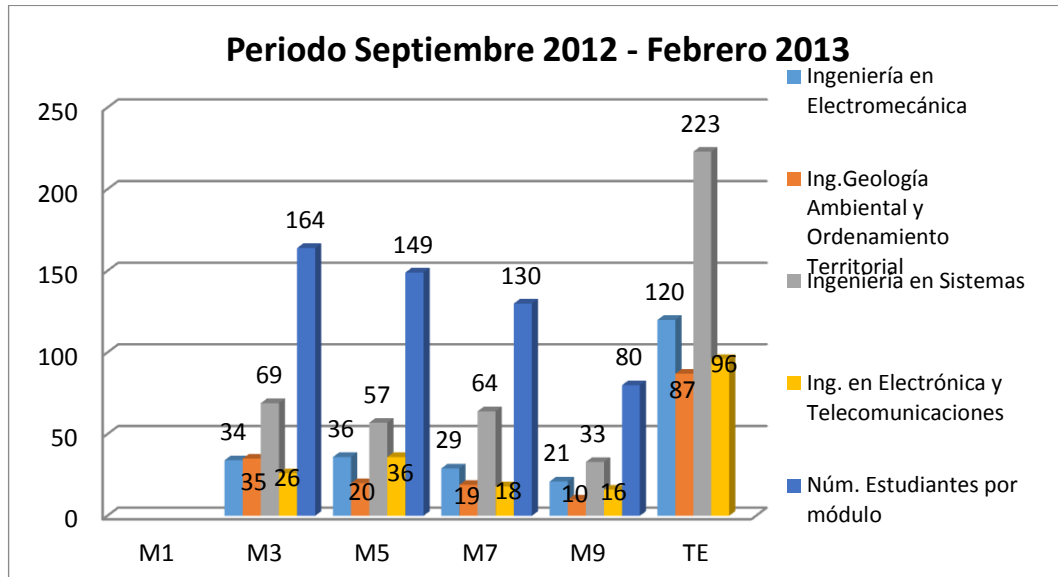


Figura 14: Estudiantes del AEIRNNR por Carrera y Módulo.

En el gráfico anterior se presenta que dentro del periodo septiembre 2012-febrero 2013 la mayor cantidad de estudiantes corresponden a la carrera de ingeniería en sistemas con 223 estudiantes, seguido de la carrera de electromecánica con 120, la carrera con menos estudiantes es la carrera de Ing. Geología Ambiental y Ordenamiento Territorial con 87 estudiantes. Además el módulo que más estudiantes tiene es el módulo 3 que son 164 estudiantes, seguido del módulo 5 con 149 estudiantes. A diferencia del módulo 9 que es el módulo con menos estudiantes con tan solo 80 alumnos por periodo académico. A partir de este periodo empieza una nueva malla curricular por lo tanto no existe el módulo 1.

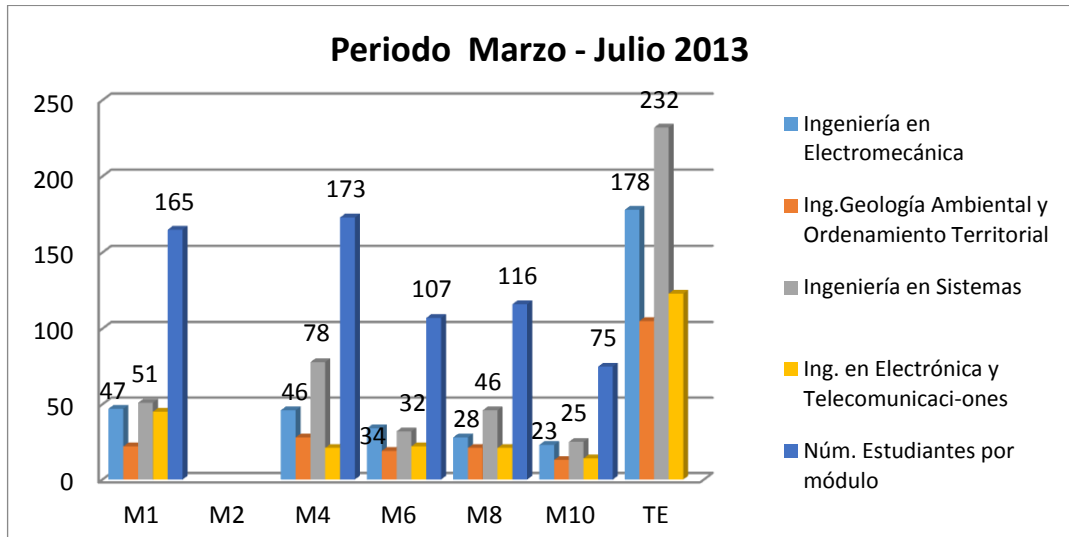


Figura 15: Estudiantes del AEIRNNR por Carrera y Módulo.

En el gráfico se presenta que dentro del periodo marzo-julio 2013 la mayor cantidad de estudiantes corresponden a la carrera de ingeniería en sistemas con 232 estudiantes, seguido de la carrera de electromecánica con 178, la carrera con menos estudiantes es la carrera de Ing. Geología Ambiental y Ordenamiento Territorial con 105 estudiantes. Además el módulo que más estudiantes tiene es el módulo 4 que son 173 estudiantes, seguido del módulo 1 con 165 estudiantes. A diferencia del módulo 10 que es el módulo con menos estudiantes con tan solo 75 alumnos por periodo académico. Debido a la ejecución de la nueva malla curricular no existe módulo 2 en todas las carreras.

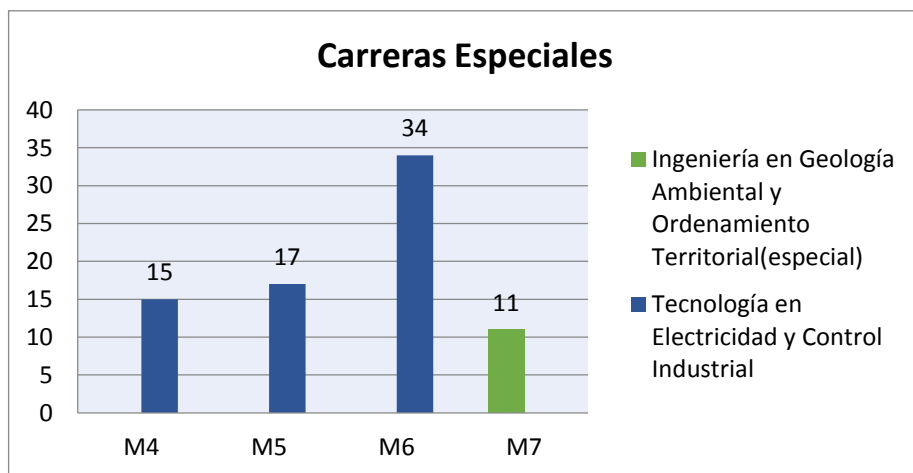


Figura 16: Estudiantes del AEIRNNR por Carrera y Módulo

Las carreras que menos estudiantes presentan son las carreras de Ingeniería en Geología Ambiental y Ordenamiento Territorial (especial) y Tecnología en Electricidad y Control Industrial, con 77 estudiantes en los 3 periodos académicos.

#### **Análisis final:**

En base al análisis realizado se puede llegar a la conclusión que la carrera con más estudiantes que evalúan por periodo académico es la carrera de Ingeniería en Sistemas, seguido de la carrera de Ingeniería en Electromecánica, a diferencia de la carrera Ing. Geología Ambiental y Ordenamiento Territorial y carreras especiales que son las carreras que presentan menos evaluaciones, en cada periodo académico. También se puede observar que los módulos con menos estudiantes en todos los periodos, son los módulos superiores como 8,9 y 10.

- **Determinación de la Muestra de Análisis.**

Para determinar la muestra de análisis, primero se realiza un estudio del número de estudiantes y docentes que pertenecen a cada módulo y carrera, dentro de cada periodo de evaluación que va desde el periodo académico 2011 hacia el periodo 2013, que pertenecen al Área de la Energía las Industrias y los recursos Naturales no Renovables. Cabe destacar que el número de estudiantes y docentes analizados, pertenecen a las carreras de Ingeniería Electromecánica, Ingeniería en Geología Ambiental y Ordenamiento Territorial, Ingeniería en Sistemas, Ingeniería en Electrónica y Telecomunicaciones, Ingeniería en Geología Ambiental y Ordenamiento Territorial (especial) y Tecnología en Electricidad y Control Industrial (Ver Tabla XXII).

Dado que el proceso de EDD se lo realiza el primer quinquemestre de estudio, se toman los estudiantes de los módulos que corresponden a un periodo de evaluación determinado. Considerando a su vez que dentro del periodo académico 2012-2013, entra en ejecución una nueva malla curricular, consideración tomada en cuenta en la respectiva tabla de análisis.

Los datos recopilados en cuanto a los estudiantes pertenecen a la base de datos actual otorgada por la UTI (Ver Anexo 5)

Para realizar operaciones dentro de las BD, es preciso conocer las clases que intervienen en el proceso de obtención de la muestra de estudiantes, a continuación las tablas descriptivas (Ver Tabla XXIII, XXIV, XXV, XXVI):

TABLA XXIII: CARRERAS AEIRNNR.

<b>Carrera</b>
Ingeniería en Electromecánica
Ingeniería en Geología Ambiental y Ordenamiento Territorial
Ingeniería en Sistemas
Ingeniería en Electrónica y Telecomunicaciones
Ingeniería en Geología Ambiental y Ordenamiento Territorial (especial)
Tecnología en Electricidad y Control Industrial

TABLA XXIV: ID PERIODOS ACADÉMICOS Y PERIODOS DE EVALUACIÓN.

<b>Id P.A</b>	<b>Periodo Académico</b>	<b>Periodo Evaluación Título Encuesta</b>
1	Septiembre 2011-Julio 2012	a. "Encuesta de Satisfacción Estudiantil 2012"
2	Septiembre 2012-Febrero 2013	b. "Evaluación Desempeño Docente 2012"
3	Marzo-Julio 2013	c. "Encuesta de Satisfacción Estudiantil 2013"

Además en cada periodo académico, existen periodos de evaluación que corresponden a una oferta académica.

TABLA XXV: ID OFERTAS ACADÉMICAS.

<b>Id</b>	<b>Oferta Académica</b>
3	Pregrado Marzo 2012 – Julio 2012
8	Pregrado Septiembre 2012 – Febrero 2013
11	Pregrado Marzo 2013 – Julio 2013

En base a la estructura relacional de la base de datos las evaluaciones en cuanto a la EDD se estructurarían de la siguiente forma:

TABLA XXVI: RELACIÓN ENTRE TABLAS DE LA EDD.

<b>Id Periodo Académico</b>	<b>Id Ofertas Académicas</b>
1	3
2	8
3	11

Se hace uso de una nomenclatura (Ver Tabla XXVII) que es una abreviación del nombre de los componentes que intervienen en el proceso de evaluación.

TABLA XXVII: NOMENCLATURA DE COMPONENTES.

<b>Nomenclatura</b>	
M	Módulo
E	Estudiantes
D	Docentes
TE	Total Estudiantes
TD	Total Docentes
PA	Periodo Académico
PE	Periodo de Evaluación

- **Muestra Estudiantes**

El reglamento de la universidad Nacional de Loja, en cuanto a la EDD determina que todos los docentes que hacen docencia de aula, serán evaluados una vez al año, terminado el primer quinquemestre de estudio. Es por eso que se limita el alcance de la presente investigación tomando las evaluaciones correspondientes desde el periodo académico 2011 al 2013. La muestra tanto de estudiantes como docentes, se establece tomando como base, los periodos de evaluación.

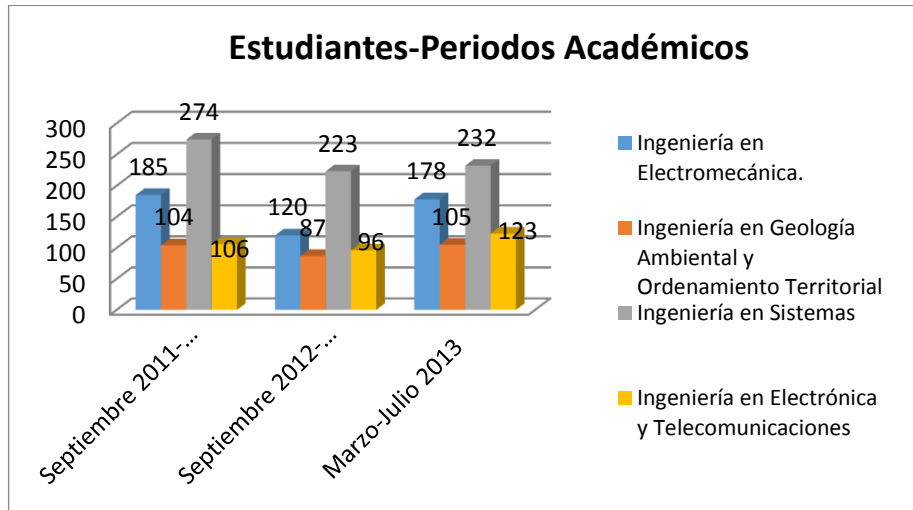


Figura 17: Número estudiantes- Periodos Académicos

En la figura anterior se puede observar que la carrera de ingeniería en sistemas, es la carrera que mas estudiantes tiene por ende más evaluaciones presenta en todos los periodos académicos con un total de 729 estudiantes, al igual que en el periodo académico septiembre 2011-Julio 2012 (Ver Figura 18), es donde existe la mayor cantidad alumnos. El resultado es un total de 1833 en los 3 periodos académicos, sumando 77 estudiantes de las carreras especiales dando un resultado final de 1910.

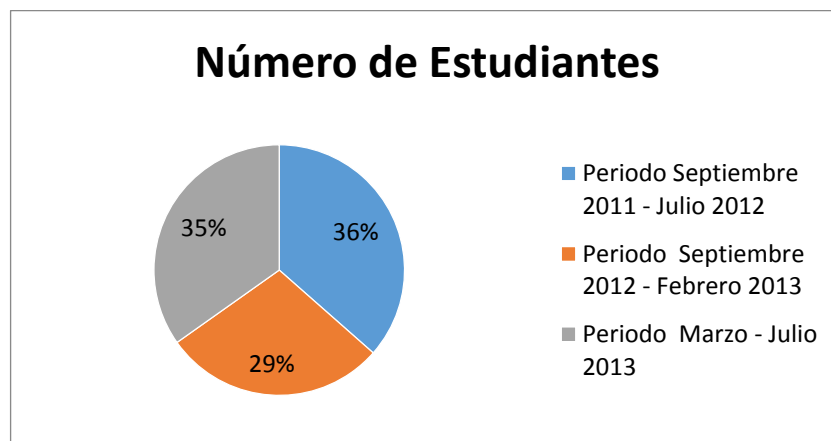


Figura 18: Número de estudiantes por periodo académico.

En el gráfico se puede observar que el periodo con mas estudiantes tiene es el periodo académico Septiembre 2011 - Julio 2012 con 669 estudiantes, y el periodo que menos estudiantes tiene es el periodo Septiembre 2012 - Febrero 2013 con 526. Por lo tanto el periodo con más evaluaciones dentro del proceso EDD es el primer periodo académico.

- **Muestra Docentes**

De la misma forma que se toma estudiantes en los periodos de evaluación, se toma el total de docentes de cada carrera. Para obtener el número de docentes en un periodo de evaluación se siguió el mismo procedimiento que con los estudiantes (Ver Tabla XXVIII, XXIX).

**TABLA XXVIII: TOTAL DOCENTES  
Docentes por Periodo de Evaluación**

<b>Carrera</b>	<b>Septiembre 2011 - Julio 2012</b>	<b>Septiembre 2012 - Febrero 2013</b>	<b>Marzo-Julio 2013</b>	<b>TD</b>
Ingeniería en Electromecánica	21	18	21	60
Ing. Geología Ambiental y Ordenamiento Territorial	12	11	14	37
Ingeniería en Sistemas	21	22	19	62
Ingeniería en Electrónica y Telecomunicaciones	12	9	11	32
<b>Total</b>	<b>66</b>	<b>60</b>	<b>65</b>	<b>191</b>

La tabla refleja que en la carrera de ingeniería en sistemas e ingeniería en electromecánica existe la mayor cantidad de docentes, pero por periodo académico tienen una similitud en número.

**Resultado del análisis**

**TABLA XXIX: TOTAL ESTUDIANTES Y DOCENTES.**

<b>Periodos Académicos</b>	<b>Total Estudiantes</b>	<b>Total Docentes</b>
2011-2012	669	66
2012-2013	526	60
2013-2013	638	65
<b>Total</b>	<b>1833</b>	<b>191</b>

La tabla refleja el total de estudiantes con 1910 estudiantes sumando 77 estudiantes de las carreras especiales y 191 docentes en los 3 periodos académicos y que a su vez corresponden a los datos dentro de un periodo de evaluación en el proceso de EDD, a



continuación se presenta la tabla total de evaluaciones recopiladas dentro de las bases de datos disponibles por periodo de evaluación (Ver Tabla XXX).

TABLA XXX: TOTAL EVALUACIONES

<b>Total de Evaluaciones por Carrera y Periodo Académico</b>									
<b>Carreras</b>	<b>2011-2012</b>			<b>2012-2013</b>			<b>2013-2014</b>		
	<b>E</b>	<b>D</b>	<b>Total</b>	<b>E</b>	<b>D</b>	<b>Total</b>	<b>E</b>	<b>D</b>	<b>Total</b>
Ingeniería en Electromecánica	185	21	3885	120	18	2160	178	21	3738
Ing. Geología Ambiental y Ordenamiento Territorial	104	12	1248	87	11	957	105	14	1470
Ingeniería en Sistemas	274	21	5754	223	22	4906	232	19	4408
In. Electrónica y Telecomunicaciones	106	12	1272	96	9	864	123	11	1353
<b>Total</b>	<b>669</b>	<b>66</b>	<b>12159</b>	<b>526</b>	<b>60</b>	<b>8887</b>	<b>638</b>	<b>65</b>	<b>10969</b>

Sumando los resultados de los 3 periodos académicos da un aproximado de 32 015 registros dentro del proceso de evaluación.

- **Periodos de Evaluación, numero de docentes evaluados**

Dentro del AEIRNNR se ha realizado el proceso de EDD en los periodos de estudio (Ver Tabla XXXI), a continuación se muestra el porcentaje de docentes que completan 3, 2 y 1 periodos académicos (Ver Figura 19).

TABLA XXXI: DOCENTES POR PERIODO ACADÉMICO

<b>Carrera</b>	<b>3 PE</b>	<b>2 PE</b>	<b>1 PE</b>
Ingeniería en Electromecánica	15	3	9
Ingeniería en Geología Ambiental y Ordenamiento Territorial	8	2	9

Ingeniería en Sistemas	9	7	21
Ingeniería en Electrónica y Telecomunicaciones	6	3	8
<b>Total docentes</b>	<b>38</b>	<b>15</b>	<b>47</b>

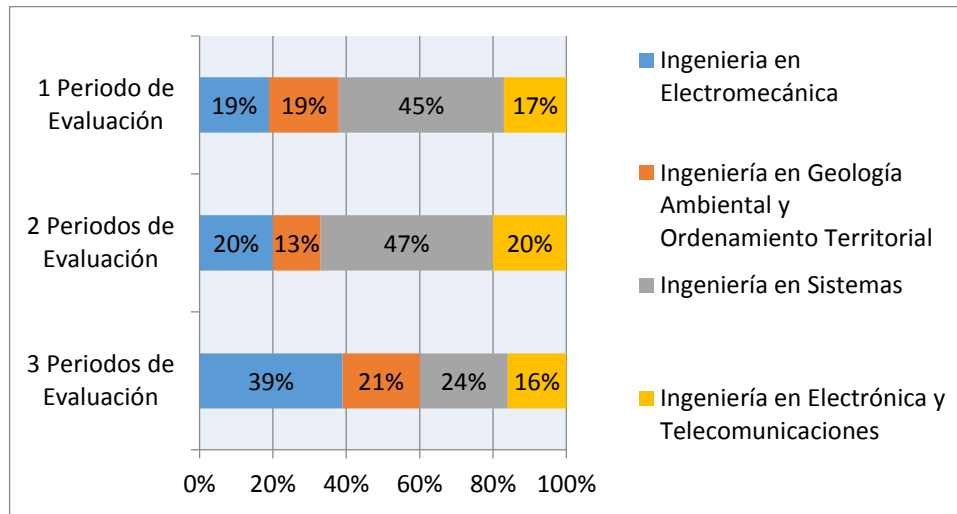


Figura 19: Docentes por periodos de evaluación

Todas las carreras tienen docentes que completan los 3 periodos de evaluación, pero en la figura se puede observar que la carrera con la mayor cantidad de docentes que completan los 3 periodos de evaluación, es la carrera de Ingeniería en Electromecánica con un 39% igual a 15 docentes, seguido de la carrera de Ingeniería en Sistemas con un 24% que es igual a 9 docentes en relación a la totalidad de docentes, que cumplen con los 3 periodos académicos y que corresponde a 38 docentes. Los docentes que completan los 2 periodos académicos son en la mayoría un 47% la carrera de ingeniería en sistemas y en la minoría la carrera de Ingeniería en Geología Ambiental y Ordenamiento Territorial. Igualmente la carrera de ingeniería en sistemas, la mayoría 45% de los docentes tienen un solo periodo académico dentro del proceso de evaluación. Se hace una excepción para las carreras especiales debido a que no completan los periodos académicos o existe una ausencia de los mismos en ciertos módulos.

A continuación se presenta el porcentaje de docentes que por carrera completan 3, 2 o 1 periodos de evaluación(Ver Tabla XXXII), (Ver Figura 20).

TABLA XXXII: DOCENTES POR CARRERA Y PERIODO ACADÉMICO

Carrera	3 PE	2 PE	1 PE	TD
Ingeniería en Electromecánica	15	3	9	27
Ingeniería en Geología Ambiental y Ordenamiento Territorial	8	2	9	19
Ingeniería en Sistemas	9	7	21	37
Ingeniería en Electrónica y Telecomunicaciones	6	3	8	17
Total 3 Periodos				100

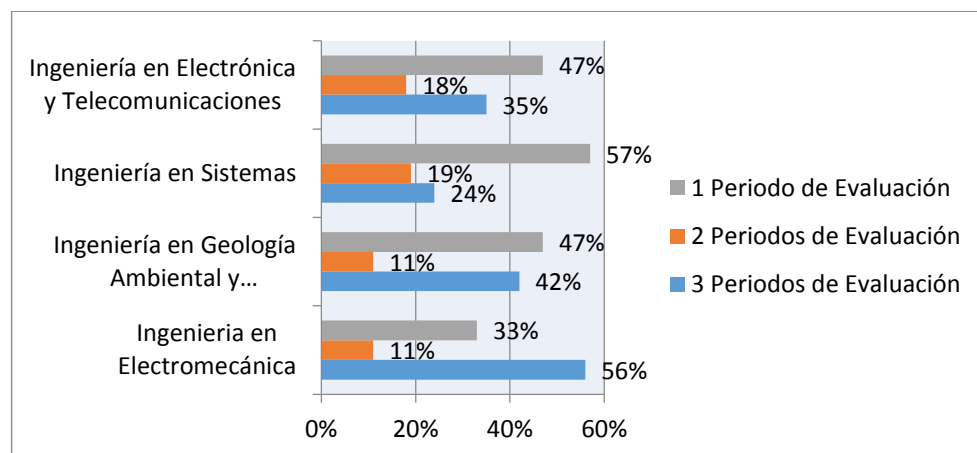


Figura 20: Docentes por periodos de evaluación

Dentro de la EDD existen docentes que ejercen la docencia completando los 3 periodos académicos, otros que están en 2 periodos y otros con 1 periodo académico. La carrera de Ingeniería en Electromecánica tiene un 56% que completan los 3 periodos, un 11% que completan 2 periodos y un 33% que tienen 1 periodo académico y por ende un periodo de evaluación. La carrera de Ingeniería en sistemas tiene su mayor cantidad de docentes que tienen un periodo de evaluación en un 57% del total de docentes en los 3 periodos que son 37, un 19% en 2 periodos y un 24% que completan los 3 periodos académicos. El total de docentes es de 100 en los 3 periodos, pero cabe señalar que este total difiere del total de docentes por periodo de evaluación.

## **1.4 Identificar la situación actual del proceso de Evaluación de Desempeño Docente.**

- **Introducción Marco Legal**

En los últimos años, muchos países han sido testigos de reformas y transformaciones en sus sistemas de educación superior, esto encierra la aparición de nuevos tipos de instituciones, cambios en los modelos de financiamiento y gobierno, además de la instauración de nuevos mecanismos de evaluación, acreditación y reformas curriculares [54, 55].

La educación superior en el Ecuador ha marcado grandes cambios en los últimos años, una nueva corriente de transformación, que va desde la creación de instituciones hasta el establecimiento de consejos como: el Consejo de Educación Superior (CES) y el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES), que aseguran primordialmente la calidad en la educación y ha provocado que vez, tras vez la educación cumpla con estándares de calidad, establecidas por el Estado [56, 57]. Es así que dentro de esta gran transformación se encuentra el proceso de evaluación al desempeño docente, que en este caso busca la calidad y excelencia de cátedra.

En cuanto a esto, la Constitución Política de la República [58], vigente en su Art. 73 dispone: “La ley regulará la carrera docente y la política salarial, garantizará la estabilidad, capacitación, promoción y justa remuneración de los educadores en todos los niveles y modalidades en base de la evaluación de su desempeño”. Es así que la evaluación tiene como objetivo principal, la obtención de información válida y confiable, con relación a la competitividad científica, ética y pedagógica que demanda una preparación universitaria. A través de la misma se ha podido evidenciar el desempeño actual de los docentes, ya que es un proceso que requiere la intervención directa de estudiantes, directivos y docentes, quienes evalúan en sí, tres parámetros, dentro de la enseñanza-aprendizaje, los cuales son: Capacidad Profesional, Capacidad Pedagógica y Práctica de Valores.

La nueva Ley Orgánica de la Educación Superior (LOES) [2], en su Art. 151, de la Evaluación Periódica Integral, dispone: “Los profesores se someterán a una evaluación

periódica integral según lo establecido en la presente Ley y Reglamento de Carrera y Escalafón del Profesor e Investigador del Sistema de Educación Superior y las normas estatutarias de cada institución del Sistema de Educación Superior, en ejercicio de su autonomía responsable. Se observará entre los parámetros de evaluación la que realicen los estudiantes a sus docentes“. Y que además “Establecerá los criterios de evaluación y las formas de participación estudiantil en dicha evaluación.”, Art 155.

Añadiendo, el Marco Reglamentario de la Universidad Nacional de Loja en su art. 25 señala [1]: “El desempeño del docente deberá ser evaluado por los estudiantes y la Comisión Académica de la Carrera o Programa, en el proceso y luego de que concluya su participación en el módulo, curso, seminario o taller”.

Las metas que persigue la evaluación del desempeño docente dentro de la Universidad:

- Convertirse en una herramienta de retroalimentación y fortalecimiento del proceso docente a fin de lograr una valoración y un mejoramiento permanente del proceso enseñanza-aprendizaje.
- Ser fuente de información a la comunidad universitaria y a su área de influencia sobre el funcionamiento de la institución respecto de la docencia, las actividades de investigación formativa y de vinculación con la sociedad que realizan los docentes en el proceso de formación profesional de los estudiantes.
- Tener una valoración objetiva de la eficacia y eficiencia del desarrollo de dichas actividades.
- Valorar al docente en la actividad académica que realiza, así como, el impacto que la capacitación y actualización permanente tienen en el mejoramiento de su desempeño; y,
- Promover un conjunto de valores institucionales que influyan en la formación profesional de los estudiantes.

Es así que este tipo de evaluaciones académicas en la educación superior se han dado en diversas universidades a nivel local, nacional e internacional [1], [59-60] en donde básicamente el objetivo primordial es mejorar el desempeño docente, es decir la

evaluación juega un papel fundamental en primer lugar para caracterizarlo y a partir de esto, buscar estrategias de mejora.

En cada una de las universidades referenciadas, la evaluación del desempeño docente por estudiantes tiene como objetivo valorar las fortalezas y dificultades de las actividades desarrolladas por los docentes en el trabajo formativo; por su naturaleza, constituye una herramienta que proporciona información útil para adecuar las estrategias institucionales de capacitación y desarrollo de competencias de los profesores, y en las cuales se evalúa: Competencia docente, Idoneidad y objetividad de la evaluación, Atención y dedicación hacia el alumno, Planificación programación del profesor, Grado de participación del alumno en clase fomentado o aceptado por el profesor, Logro de objetivos y metas y Percepción global del desempeño docente.

Es necesario tener presente que la Evaluación del Desempeño Docente, en la Universidad Nacional de Loja, además de ser un requerimiento institucional, ligado al reglamento de educación superior, es un camino u oportunidad de mejoramiento en el campo científico-técnico, profesional y pedagógico, del docente, por lo que es necesario, tomar en cuenta cada aspecto, o parámetro que interviene en el proceso y resultado de la misma. Por tal motivo todos los actores involucrados, se ven enmarcados en un accionar serio, válido y confiable, para evaluar y valorar el desempeño continuo de los docentes, con la perspectiva de verificar, retroalimentar y mejorar la calidad de los procesos de aprendizaje y enseñanza en la Universidad.

- **Evaluación del desempeño docente.**

Al referirnos a la evaluación del desempeño docente, se puede decir que son un conjunto de acciones organizadas en base a las intervenciones e interacciones pedagógicas, entre el conocimiento científico y el conocimiento académico, lo cual permitirá impulsar acciones didácticas pedagógicas que fortalezcan los procesos de aprendizaje de los estudiantes, y el mejoramiento de la formación inicial docente, y a su vez su desarrollo profesional [61].

Además la evaluación del desempeño docente es un proceso sistemático de obtención de información confiable, con el objetivo de comprobar y valorar el efecto educativo que

produce en los estudiantes, el despliegue de sus capacidades en la docencia, investigación, vinculación con la colectividad, en la cooperación interuniversitaria y compromiso institucional [1].

A nivel institucional, se ha entendido y puesto en práctica una evaluación interna sobre el desempeño docente, en las que se evalúan principalmente las competencias y actitudes que el profesional de la educación pone en práctica en la institución educativa y en los procesos de formación y aprendizaje de los estudiantes, que contribuyen al cumplimiento de los objetivos educativos e institucionales. La evaluación del desempeño docente juega un papel de primer orden que permite caracterizar al docente, y por tanto, propiciar su desarrollo futuro. La evaluación del docente es una forma de fomentar y favorecer su permanente perfeccionamiento.

En la actualidad, hay un consenso de que el fracaso o el éxito de todo proceso educativo, depende fundamentalmente de la calidad del desempeño docente. En una ponencia de Cuba en el Encuentro Iberoamericano sobre Evaluación del Desempeño Docente, se sostuvo que: “Podrán perfeccionarse los planes de estudio, programas, textos; construirse magníficas instalaciones; obtenerse excelentes medios de enseñanza, pero sin docentes eficientes no podrá tener lugar el perfeccionamiento real de la educación” [7] o dicho en otras palabras una educación de calidad.

- **Calidad.**

Al hablar de calidad, se percibe rápidamente una idea de participar de lo mejor. Añadiendo, según la Norma INTECO ISO 900 [62], la calidad es: Conjunto de propiedades y característica de un producto o servicio, que le confiere su aptitud para satisfacer necesidades expresadas o implícitas. En cuanto a procesos educativos, se puede entrever que el objetivo principal de la calidad es un buen rendimiento o un mejoramiento continuo en base a los recursos académicos disponibles.

En la Ley de Educación Superior Art. 93, se nombra un Principio de Calidad que consiste en la búsqueda constante y sistemática de la excelencia, la pertinencia, producción óptima, transmisión del conocimiento y el desarrollo del pensamiento mediante la autocrítica, la crítica externa y el mejoramiento permanente [2].

Pero además el Estado mismo garantiza, a través de la Constitución Art 27 [58], el cumplimiento y ejercicio de la calidad como factor primordial de la educación.

La calidad es el fin que todos persiguen independientemente del área de servicio o sistema que rige cualquier institución, es decir la calidad es la influencia pasada, presente y futura para el éxito de cualquier proyecto y en esta está vinculada intrínsecamente la educación.

- **Evaluación de la calidad en la educación.**

- **Reseña Histórica.**

La evaluación de la calidad en la educación se ha dado desde tiempos antiguos por ejemplo, a inicios del siglo XX, en el marco del desarrollo industrial de los Estados Unidos y en el afán de adaptar la educación al contexto de la calidad, se suscita una nueva aproximación a la evaluación sistemática del aprendizaje, sustentada en la psicología científica y la lógica de la administración del trabajo. El educador americano Abraham Flexner hizo una evaluación de las escuelas de medicina de los Estados Unidos y Canadá en 1910 y llegó a la conclusión de que 120 facultades (de las 155 analizadas), mostraban pésimas condiciones de funcionamiento.

Como resultado de esta evaluación, casi todas las instituciones que obtuvieron una mala calificación por parte de Flexner, cerraron sus puertas [63].

Desde entonces, la evaluación realizada por Flexner se ha convertido en un ejemplo de la importancia de la evaluación y del control de calidad de los procesos educativos. Por ello, tanto su actuación como su concepto de calidad de servicio tienen vigencia hasta nuestros días.

Más adelante y con el auge tecnológico se vinieron a establecer las pruebas objetivas como, instrumentos que permitirían la medición de lo aprendido en un proceso educativo. Siendo así, la calidad uno de los ámbitos más estudiados en la evaluación de la educación [64].



Ya en la década de los sesenta, aparecen en América Latina modelos y estudios institucionales con el apoyo conceptual de la teoría de sistemas.

Sustentada por Scriven [65], la teoría de sistemas diferencia entre evaluación formativa y evaluación sumativa.

La evaluación formativa se realiza durante el desarrollo del programa o plan de estudios y produce información útil para mejorar su desenvolvimiento, en cambio la evaluación sumativa se efectúa al finalizar el programa de estudios, con el propósito de determinar el valor del programa para los usuarios o en este caso los educandos.

La evaluación formativa permitió dirigir la vista hacia los procesos educativos y no sólo hacia los productos de éstos. Siendo a su vez una llave que permitió la entrada a nuevas maneras de concebir el problema de la evaluación.

Hacia finales de la década de los sesenta, se comienza a vincular la evaluación con la toma de decisiones, y es a partir de entonces, se observa un interés mayor por abarcar todos los procesos educativos y aun los sistemas educativos, desde la óptica de la evaluación.

En la actualidad en nuestro país la evaluación de la calidad ha sido un factor determinante dentro de la educación, de tal manera que la LOES Art. 94 [2], menciona que “La Evaluación de la Calidad es el proceso para determinar las condiciones de la institución, carrera o programa académico, mediante la recopilación sistemática de datos cuantitativos y cualitativos que permitan emitir un juicio o diagnóstico, analizando sus componentes, funciones, procesos, a fin de que sus resultados sirvan para reformar y mejorar el programa de estudios carrera o institución.”

Un plan de evaluación debe buscar la calidad de la educación superior, evaluando a los docentes para tomar los correctivos que sean necesarios y también propender a que la institución califique satisfactoriamente con la evaluación Institucional sobre la calidad educativa Art.9 Reglamento (LOES) [2], partiendo con una autoevaluación interna que determine las necesidades de cambios para posteriormente ser evaluados por el Consejo de evaluación, acreditación y aseguramiento de la calidad de la Educación Superior (CEAACES).

A nivel local la UNL, en su proyecto de evaluación de desempeño docente, asume el compromiso de responder con calidad, pertinencia y eficiencia, las actividades llevadas a cabo dentro del proceso, no solo por un mandato constitucional, sino por convicción de asegurar su calidad a través de la evaluación, que a su vez facilitará la rendición social de cuentas [1].

- **Identificación de Problemas actuales.**

El proceso de evaluación al desempeño docente dentro de la Universidad Nacional de Loja, en la actualidad, tiene resultados, que miden el desempeño en base a las puntuaciones obtenidas en el rendimiento de los docentes, por medio de las encuestas realizadas tanto a la comisión académica de cada área y a los estudiantes que culminan cierto módulo. Las calificaciones o resultados van en un rango de: en insatisfactorio a muy satisfactorio, pero no existe un mecanismo exacto que evidencie la opinión de los alumnos en los informes finales, dirigidos hacia el docente evaluado, y con esto tener mayor sustento en cuanto a la elaboración del plan de mejoras al desempeño profesional del docente.

Si bien es cierto el proceso de evaluación docente es analizado, basado en varios puntos muy acertados; como el desempeño, en base a la Capacidad Pedagógica, Capacidad Profesional y La Práctica de Valores [1], los criterios emitidos por cada uno de los estudiantes dentro de este proceso deberían ser analizados más a profundidad. Ya que actualmente los comentarios y sugerencias de los estudiantes solo reafirman la respuesta solicitada por el cuestionario. De tal forma que a través de un análisis minucioso de los criterios estudiantiles se pueda dar el descubrimiento de conocimiento o detección de anomalías que existe dentro de las Bases de Datos que corresponden a este proceso.

Con esto y mediante la observación se puede evidenciar el poco control de la gestión del docente (en cuanto al plan o planes de mejora) y la ausencia de un método sistemático de recopilación de datos que sustente la evaluación.

Esta la importancia o el impacto que tienen los criterios de los estudiantes socializados hacia el docente; si el profesor tiene un conocimiento de sus fortalezas podrá acrecentarlas y si conoce sus debilidades podrá corregir sus fallas, lo que sólo podrá

lograr si cuenta con una evaluación completa y objetiva del desempeño de su labor docente [20].

En el AEIRNRR, dicha evaluación completa si se realiza, pero no se socializa de forma completa todos los parámetros que intervienen en este proceso. Siendo los comentarios y sugerencias de los estudiantes un ente muy importante dentro del proceso enseñanza-aprendizaje del área.

Conforme se mejore el proceso enseñanza-aprendizaje se propicia en el alumno su independencia, el logro de los objetivos educativos propuestos, favorece la adquisición de habilidades de estudio y trabajo autónomo. Como consecuencia lógica de una mejora en los métodos de trabajo del profesor, es posible esperar también una reducción en los índices de deserción y rezago, así como un incremento en el aprovechamiento académico y una mejora en los índices de titulación.

Estudios recientes revelan que uno de los problemas que enfrentan las instituciones de educación superior a nivel de Latinoamérica es el bajo aprovechamiento académico, la deserción, el rezago estudiantil y los bajos índices de eficiencia terminal [66-68]. Y la educación superior local, no es la excepción dentro de estos problemas.

Si nos basamos en un análisis comparativo de periodos académicos anteriores, podemos decir que si bien es cierto, que el proceso de evaluación de desempeño docente ha pretendido evaluar los aspectos cognitivo, pedagógico y valores de los docentes, esta puede con llevar a un sesgo de la información, dando menos importancia al juicio valorativo de los estudiantes, dentro del proceso, además no se han implementado mecanismos de seguimiento por procesos que abarquen íntegramente la gestión del docente en todas sus actividades académicas desarrolladas en el aula, así como tampoco se han conformado comisiones especiales para evaluar presencialmente el desempeño docente.

En este sentido, muchos autores han afirmado que la calidad de los docentes es uno de los factores más importantes al interior de una institución educativa para predecir el desempeño de los estudiantes. Es decir que, los estudiantes que han tenido docentes de

alta calidad logran un nivel de aprendizaje tres veces mayor que los estudiantes que tienen docentes de baja calidad [69, 70].

En cuanto a los docentes, una evaluación que apunte a los estándares mínimos están diseñados para que entren o se mantengan profesores de bajo rendimiento; los estándares competitivos tienen como propósito proporcionar incentivos y compensaciones por alguna clase de rendimiento superior; y, los estándares de desarrollo están diseñados para fomentar el desarrollo de los conocimientos y capacidades de los profesores especificando las áreas de capacidades u objetivos de mejora [60, 70]. Ninguno de estos estándares se cumple actualmente dentro del AEIRNNR.

Como conclusión vale la pena destacar que, la evaluación del desempeño docente, a más de ser un mandato legal, se constituye en el derecho que tienen los docentes a mejorar su práctica profesional, pero también a beneficiarse mediante sus logros, en acceso a incentivos y compensaciones a partir de un proceso de evaluación serio, transparente y confiable.

## **2. FASE 2: Analizar las técnicas de Minería de Datos (MD) factibles para el proceso de Evaluación al Desempeño Docente**

### **2.1 Búsqueda de información en fuentes académicas, revistas, tesis doctorales y artículos científicos sobre las diversas técnicas de MD.**

La Minería de Datos es un proceso que consiste en analizar los datos desde diferentes perspectivas, en base a la extracción de los mismos y convirtiéndolos en información útil. Pero además permite descubrir las relaciones en bases de datos que pueden identificar comportamientos que no son muy evidentes, es decir permite la identificación de patrones ocultos [71, 72].

Además las técnicas de MD incorporan diferentes algoritmos (Ver Tabla XXXIII) que permiten la generación de modelos [73]. En la actualidad diversas investigaciones orientan sus proyectos a la aplicación de Data Mining como opción o solución a problemas empresariales como institucionales, obteniendo muy buenos resultados [19-21], [74, 75]. Cabe mencionar que toda la información procede de varias fuentes bibliográficas las mismas que están descritas dentro de la Revisión Literaria en el Capítulo II.

**TABLA XXXIII: RECOPIACIÓN TÉCNICAS DE MINERÍA DE DATOS**

<b>Técnicas</b>	<b>Descripción</b>
<b>1. Árboles de Decisión</b>	<ul style="list-style-type: none"><li>• Los árboles de decisión han sido motivo de gran estudio en el ámbito de la minería de datos [76-78], estos son conjunto de reglas representadas en forma de una estructura de árbol las cuales son muy útiles en cuanto a la representación del conocimiento en una base de datos.</li><li>• Entre sus ventajas se tiene que son fáciles de entender, no se requieren suposiciones acerca de los datos y maneja variables "input" que pueden ser continuas y categóricas.</li><li>• Se puede decir que existen muchos algoritmos para construcción de árboles de decisión pero que lo que</li></ul>

	<p>los diferencia virtualmente a todos es la regla usada para la disgregación del árbol y la estrategia de poda.</p>
<p><b>2. Redes Bayesianas</b></p>	<ul style="list-style-type: none"> <li>• Formalmente, una Red Bayesiana es un grafo dirigido a cíclico cuyos nodos representan variables y los arcos que los unen codifican dependencias condicionales entre las variables. El grafo proporciona una forma intuitiva de describir las dependencias del modelo y define una factorización sencilla de la distribución de probabilidad conjunta consiguiendo un modelo manejable que es compatible con las dependencias codificadas.</li> <li>• En particular las Redes Bayesianas constituyen una metodología elegante y potente basada en la probabilidad y la estadística que permite construir modelos de probabilidad conjunta manejables que representan las dependencias relevantes entre un conjunto de variables.</li> <li>• Los modelos resultantes permiten realizar inferencia probabilística de una manera eficiente.</li> <li>• Las redes bayesianas también se han establecido como una técnica de minería de datos en diferentes estudios e investigaciones [79-81].</li> </ul>
<p><b>3. Redes neuronales</b></p>	<ul style="list-style-type: none"> <li>• Las redes neuronales se han revelado como un útil instrumento para obtener información a partir de grandes masas de datos [82-84].</li> <li>• Las redes neuronales son un método de aprendizaje que fue inspirado de la forma en que funciona el sistema nervioso de los animales.</li> <li>• Principalmente las Redes Neuronales Artificiales (RNA), son útiles en problemas de predicción, reconocimiento de patrones, el agrupamiento, la clasificación, la visualización, etc. [85].</li> </ul>

<p><b>4. Regresión Logística</b></p>	<ul style="list-style-type: none"> <li>• La regresión logística es un tipo especial de regresión que se utiliza para predecir y explicar una variable dependiente.</li> <li>• Esta técnica estadística también se considera como una técnica de análisis discriminante que permite explicar alguna característica o analizar problemas con una variable independiente o la combinación de varias variables independientes.</li> <li>• Esta técnica se interesa en la predicción y explicación de las relaciones que influyen en la categoría en que un objeto está situado.</li> <li>• Es decir en este modelo, el problema es encontrar la estimación de la probabilidad de ocurrencia de un evento, dado que ha ocurrido otro. En otras palabras, la variable respuesta representa la probabilidad condicional de ocurrencia de un evento.</li> <li>• Una ventaja adicional de esta técnica es que no requiere la normalidad estricta de los datos, además muchos estudios han evidenciado otras características que hacen de la regresión logística una buena herramienta para la categorización [86-89].</li> </ul>
<p><b>5. Agrupamiento o Clustering</b></p>	<ul style="list-style-type: none"> <li>• Las técnicas de clustering son técnicas de clasificación no supervisada de patrones en conjuntos denominados clústers [90, 91].</li> <li>• El problema del clustering ha sido abordado por gran cantidad de disciplinas y es aplicable a una gran cantidad de contextos, lo cual refleja su utilidad como uno de los pasos en el análisis experimental de datos [90, 91].</li> <li>• Las técnicas de clustering han sido ampliamente utilizadas en múltiples aplicaciones y sirve para descubrir patrones de distribución general y</li> </ul>

	<p>correlaciones interesantes entre los atributos de los datos [92].</p> <ul style="list-style-type: none"> <li>• El análisis de clústers se puede usar para hacerse una idea de la distribución de los datos, para observar las características de cada clúster y para centrarse en un conjunto particular de datos para futuros análisis.</li> <li>• A diferencia de la clasificación, el clustering no depende de clases previamente definidas ni en ejemplos de entrenamientos etiquetados a priori.</li> </ul>
<p><b>6. Reglas de Asociación</b></p>	<ul style="list-style-type: none"> <li>• Las reglas de asociación también son una de las técnicas más utilizadas dentro de la minería de datos, obteniendo reglas y asociaciones que describen el suceso u ocurrencia de un evento.</li> <li>• Las reglas que cumplan con un alto soporte y confidencialidad en las asociaciones generadas serán las que serán tomadas en cuenta dentro de cualquier investigación [34].</li> </ul>

## **2.2 Investigar y analizar herramientas útiles en el proceso, y aplicación de técnicas de MD.**

El análisis de las herramientas constituyó un punto indispensable en la presente investigación ya que las mismas ayudaron en el ambiente de trabajo proporcionando la ayuda en la ejecución de cada tarea.

### **2.2.1 Herramientas útiles en el proceso de Minería de Datos**

En este punto se procedió a analizar (Ver Anexo 6) y evaluar herramientas útiles en el proceso de minería, en base a las características y beneficios que estas presentan (Ver Tabla XXXIV) [14-25]).



TABLA XXXIV: CARACTERÍSTICAS DE HERRAMIENTAS DE MINERÍA DE DATOS.

Herramienta	Metodología	GUI	Lic.	Manipulación a través de:			Integración con Herramientas
				Línea de Comandos	Creando Aplicación	Batch (Lotes)	
	Varias	si	GNU	si	no	no	no
	Varias	si	GNU	si	si	no	si
	Varias	si	GNU GPL	si	si	si	si

La tabla presenta las características de cada herramienta evaluada en cuanto al proceso de obtención de conocimiento cabe mencionar que la descripción de las mismas se elaboró en el Capítulo III de Revisión Literaria.

### 2.2.1.1 Evaluación final

Luego de haber investigado y analizado los servicios y privilegios con que cuentan las herramientas presentadas (Ver Tabla XXXV) se observó que:

TABLA XXXV: CARACTERÍSTICAS PRINCIPALES HERRAMIENTAS MD - EVALUACIÓN FINAL

Consideraciones principales de cada herramienta		
<b>Knime:</b> ofrece a los usuarios la capacidad de crear de forma visual flujos o tuberías de datos, ejecutar selectivamente algunos o todos los pasos de análisis, y luego estudiar los resultados, modelos y vistas interactivas. Además su uso	<b>Weka:</b> Está constituida por una serie de paquetes de código abierto con diferentes técnicas de pre-procesado, clasificación, agrupamiento, asociación y visualización. Este programa se distribuye como software de libre	<b>Rapidminer:</b> permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. Además, esta aplicación ofrece más de 500 operadores para todos los principales

<p>se basa en el diseño de un flujo de ejecución que plasma las distintas etapas de un proyecto de minería de datos.</p>	<p>distribución (licencia GNU-GPL, además permite la implementación de algoritmos para el pre-procesamiento de datos. Trabaja con archivos de formato arff y permite además la opción de acceso a bases de datos sql.</p>	<p>procedimientos de máquina de aprendizaje, y también combina esquemas de aprendizaje y evaluadores de atributos del entorno de aprendizaje Weka.</p>
--	---	--





De la mano de las herramientas útiles en la minería, se encuentran las herramientas para la gestión de base de datos las cuales permiten la recopilación, tratamiento, y estructuración de los datos para su posterior uso.

### **2.2.2 Herramientas, útiles para gestión de Bases de Datos.**

Para la preparación de los datos y en cuanto a la generación de estructuras se analizó los gestores de base de datos convenientes a los datos disponibles.

La siguiente tabla (Ver Tabla XXXVI [93-96]) tiene una breve descripción de gestores de base de datos (BD), en donde se tomó algunas herramientas, considerado sus características como el formato de importación y exportación para su elección (Ver Figura 21).

TABLA XXXVI: CARACTERÍSTICAS DE HERRAMIENTAS PARA ADMINISTRAR BD.

Herramienta	Formato de importación			Formato de Exportación							Asistente: Crear/Modifi car Datos	Licencia
	CSV	ZIP	SQL	CSV	DBF	TXT	XLS	XML	HTML	DIF		
 pgAdmin III	si	si	si	si	si	si	si	no	no	si	si	Free [93]
 SQLyog Enterprise	si	no	si	si	no	no	no	si	si	no	si	Trial [94]
 DataAdmin 6.4.2.5 Personal	no	si	si	si	no	si	si	si	si	no	si	Personal[95]
 phpMyAdmin	si	no	si	si	no	si	si	si	no	no	si	Free [96]

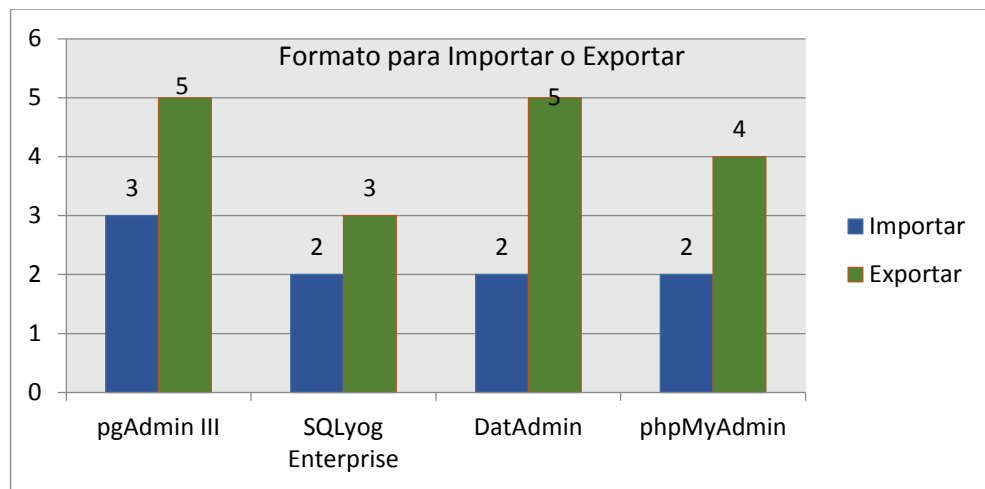


Figura 21: Formatos de archivos por cada herramienta.

**Análisis:** la herramienta con menos posibilidades en cuanto a formatos presenta es la herramienta SQLyog Enterprise, a diferencia de DatAdmin que presenta el mayor formato de exportación, en cuanto a la licencia, cuenta con las características apropiadas para la manipulación de los datos. PhpMyAdmin sigue en cuanto a ventajas con un 4 al exportar archivos y un 2 al importarlos sería una buena elección luego de pgAdmin III ya que la misma tiene 5 formatos de los archivos analizados para exportar y 3 formatos para importar.

Cabe recalcar que existen diversas herramientas para la administración de BD que presentan más o diferentes características de uso.

### 2.2.3 Herramientas, de apoyo al proceso de Minería de Datos

Existen diversas herramientas que sirven de apoyo al proceso de minería (Ver Anexo 6) ya sea para la interpretación de resultados, visualización de gráficos entre otros.

A continuación se presenta una breve descripción de las mismas (Ver Tabla XXXVII).

TABLA XXXVII: HERRAMIENTAS DE APOYO AL PROCESO DE MINERÍA

<b>Herramienta</b>	<b>Descripción</b>
<b>Many Eyes</b>	Many Eyes es un conjunto de herramientas gratuitas de visualización de datos proporcionados por IBM. El sitio permite a los usuarios cargar los datos y producir sus visualizaciones de manera que otros usuarios puedan verlas y comentarlas. Los usuarios tienen una amplia gama de opciones de visualización disponibles – diagramas de red, diagramas de dispersión, diagramas matriciales, gráficos de barras, histogramas, gráficos de burbujas, gráficos de líneas, gráficos apilados, gráficos circulares, mapas árbol, árboles de palabras, nubes de etiquetas, generador de palabra nube, los así como los mapas del mundo y EE.UU. [97].
<b>Gephi</b>	Gephi es un desarrollo de código abierto, multiplataforma bajo la licencia GNU GPL, es una de las herramientas más completas

	<p>para manejar todo tipo de datos y gráficos con un buen desempeño, incluso al manejar y procesar datos sumamente grandes.</p> <p>Gephi es útil para los analistas de datos a hacer hipótesis, aislar estructuras de datos o descubrir patrones de una forma intuitiva, todo con la ayuda de una serie de características interesantes, tales como el análisis dinámico de datos; creación de cartografía con el fin de poder visualizar solo un conjunto de datos que en realidad sea de interés, modificando el tamaño de los nodos o resaltando la importancia de algunos de los enlaces y aislar los datos pertinentes o el uso de métricas preestablecidas [98].</p>
<p><b>Textalyser</b></p>	<p>Es un programa gratuito destinado a analizar la complejidad de textos y a aportar estadísticas sobre la frecuencia de palabras. Lleva a cabo análisis en archivos originales en PDF o en cualquier otro archivo de texto se puede obtener un informe de las palabras y frases utilizadas en el mismo. También se puede cargar directamente el archivo. Da como resultado un informe con las frecuencias de frases, palabras y otros indicadores de complejidad textual, además da una relación de las palabras individuales y su frecuencia. [99]</p>

La tabla describe cada una de las herramientas útiles en un análisis gráfico, presentadas las características principales de cada una, se optó por la herramienta Many Eyes, ya que han sido con anterioridad útiles en procesos de minería de textos [99, 100] y es de acceso libre.

## **2.3 Selección de las herramientas y técnicas para el análisis de evaluación docente.**

Luego del análisis de las herramientas tanto para gestión de Bases de Datos, como para los procesos de minería y visualización de gráficos descritos anteriormente se optó por las siguientes herramientas y técnicas:

### **2.3.1 Selección de Herramientas**

- Para el presente trabajo se eligió un SGBD que permita principalmente trabajar con datos en Postgres, por ello se eligió trabajar con PgAdmin III, debido, a que cuenta con más ventajas en cuanto a la importación y exportación de los datos, cuenta también con una licencia libre que permite una manipulación de datos satisfactoria y porque además tiene una interfaz gráfica muy amigable que permite su facilidad de uso.

Por lo tanto la gestión de la BD proporcionada por el SGA, se la hizo en la herramienta pgAdmin III, que en base al presente análisis, es la más adecuada en cuanto a características de uso.

- En cuanto al análisis de herramientas, útiles en el proceso de minería de datos, se ha determinado que Weka y RapidMiner son las herramientas que ofrecen más ventajas para el trabajo que se desea llevar a cabo, ya que ambas herramientas se complementan. Pero se optó por la herramienta RapidMiner por la variedad de características que ofrece en cuanto al proceso de minería de datos, y porque en Rapidminer se incluyen muchas funcionalidades de la herramienta Weka. A su vez Weka ayudó en la aplicación de algoritmos que en RapidMiner no están bien definidos.
- En cuanto a la visualización de gráficos y aporte al tex-mining se eligió la herramienta Many Eyes debido a que ha sido útil en procesos de minería de textos [99, 100] y es de acceso libre.

### 2.3.2 Selección de Técnicas

Para la selección de los algoritmos se apoyó en los resultados obtenidos en cada caso de éxito revisado anteriormente (Ver Tabla XXXVIII). Ya que los mismos dieron una pauta de las técnicas y herramientas utilizadas para resolver determinado problema.

TABLA XXXVIII: SELECCIÓN TÉCNICAS DE MINERÍA DE DATOS

<b>Caso de Éxito</b>	<b>Técnicas</b>	<b>Algoritmos</b>	<b>Herramientas</b>
Minería de Datos en la Educación	Agrupamiento-Clúster, Árbol de Decisión, Generación de Reglas, Regresión, Aprendizaje Inductivo.	K-means, CHAID, ID3, Reglas asociación, Regresión logística, Regresión Lineal	Answer Tree, SPSS, Weka, Rapid miner
Minería de Datos en Encuestas de Profesores al Fin de Semestre de la Facultad de Ingeniería, UNAM.	Agrupamiento-Clúster, árboles de decisión, redes neuronales	K-means, Reglas asociación	SPSS, Weka, Rapidminer
Obtención de Patrones y Reglas en el Proceso Académico de La Universidad de Las Ciencias Informáticas Utilizando Técnicas de Minería de Datos.	Agrupamiento-Clúster, Generación de Reglas Regresión.	Árboles de decisión, Regresión Logística	SQLServer, Google Refine, Rapidminer, CRISP-DM

La tabla anterior recopila cada caso de éxito abordado dentro de la Revisión Literaria, esto proporcionó una pauta para la elección de las técnicas para trabajar en la presente investigación, por lo tanto se eligió las técnicas de clasificación como árboles de decisión, técnicas de regresión, generación de reglas y técnicas bayesianas.

### **3. FASE 3: PROCESO DE MINERÍA DE DATOS**

#### **3.1 Procesos y Tratamiento de los datos en la elaboración de un modelo de minería de datos**

En los siguientes puntos se fue desglosando el proceso de minería de datos esto va desde la comprensión del negocio hasta el modelado.

##### **3.1.1 Comprensión del Negocio o Problema**

Etapa inicial de la investigación que consistió en determinar y definir claramente los problemas a los que se va a dar solución.

###### **3.1.1.1 Determinar los objetivos del negocio**

Los objetivos que se desearon alcanzar con el desarrollo del presente trabajo de titulación, tiene su base en la idea de analizar el proceso de Evaluación de Desempeño Docente (EDD) aplicando técnicas de Minería de Datos.

La Universidad Nacional de Loja, con la visión de las autoridades y con la finalidad de certificar una educación de calidad, pero sobre todo de asegurar la eficacia en la enseñanza en la formación profesional de los estudiantes, ha implementado el proceso de evaluación del desempeño docente.

El problema radica que actualmente el proceso de evaluación al desempeño docente no cuenta con mecanismos que filtren la información en base a los criterios de los estudiantes dentro del proceso, o describan que factores inciden en la calificación de un docente y menos aun no indican si un docente ha tenido o no una evolución de cátedra.

El criterio u observaciones descritas por estudiantes dentro de la evaluación, actualmente deben ser analizados de forma individual en base al docente evaluado. Dando como consecuencia un proceso largo y tedioso que el usuario o administrador encargado de emitir los reportes finales, debe asumir. Considerando además que la cantidad de docentes y estudiantes dentro de la universidad en todas sus extensiones es alta; por tal motivo el proceso de verificación de información puede durar hasta meses.



Luego de haber dado una idea general del proyecto, a continuación se describen los objetivos dentro del marco de la comprensión del negocio y los criterios de éxito del proyecto:

- **Objetivos del negocio**

- Obtener los criterios de interés (opiniones de estudiantes) y buen desempeño docente, que se puedan dar dentro del proceso.
- Observar la evolución de cátedra que han tenido los docentes en los periodos académicos tomados para el estudio.
- Determinar los patrones de comportamiento de estudiantes que evalúan a sus docentes luego de obtener la nota dentro de un módulo y visualizar la evaluación en base a calificaciones.

- **Criterios de éxito**

- La identificación de patrones dentro del proceso de evaluación es tomado como un criterio de éxito en el desarrollo del presente trabajo investigativo, tomando como base los datos de EDD del AEIRNNR de la Universidad Nacional de Loja., en donde se llevó a cabo el tratamiento respectivo de los mismos.
- Las notas de los estudiantes también se utilizan, para detectar el comportamiento de los estudiantes al momento de evaluar, y la evolución de cátedra que un docente tiene en base a las evaluaciones, fueron criterios de éxito dentro de la minería de datos en el presente proyecto.

### **3.1.1.2 Evaluación de la Situación**

La información que se obtuvo acerca de la problemática que se planteó resolver fue la adecuada, precisa y suficiente, ya que además de tener acceso a los datos se mantuvo entrevistas con el encargado de la UTI y la coordinadora de EDD, lo cual posibilita la ejecución y conclusión satisfactoria del presente trabajo de titulación.

Dentro de la evaluación de la situación se llevó a cabo un análisis de recursos disponibles para el desarrollo de la investigación, recursos que abarcan todo el medio de ejecución del proyecto como talento humano, hardware, software, disponibilidad de datos o información, materiales etc. Además se detallan los riesgos y contingencias que pueden repercutir o demorar el proyecto.

### **3.1.1.2.1 Recursos Disponibles**

Los recursos disponibles fueron la plataforma en la ejecución del proyecto, la disponibilidad y acceso a los mismos permitió el cumplimiento de cada una de las fases de desarrollo planteadas dentro del proyecto.

#### **3.1.1.2.1.1 Recursos Humanos**

En el presente proyecto el coste del personal se entiende por el número de horas en la realización de las actividades del proyecto. Los trabajadores que pertenecen al equipo de trabajo son los siguientes:

- **Director de Proyecto:** su función consistió en la revisión constante del desarrollo del proyecto, con el fin de obtener coherencia e idoneidad de las actividades, así como supervisión y guía en la planificación, adecuación, e impacto de la temática de la investigación.
- **Investigador:** las actividades realizadas fueron investigación, manejo de información, recolección de datos, administración de herramientas tanto para la gestión de datos como para el proceso de minería de datos, uso de técnicas y algoritmos en el desarrollo y elaboración del modelo de Data Mining.
- **Administrador UTI:** persona encargada de proporcionar el acceso a los datos, creando además un servicio de web services adicional al que existe, con el fin de obtener las notas de los docentes.

#### **3.1.1.2.1.2 Recursos Hardware:**

Para el desarrollo del proyecto se utilizó recursos hardware como computador personal, dispositivos de almacenamiento externo como flash memory e impresora.

- **Computador personal:** en el desarrollo del proyecto se hizo uso de un computador de escritorio con características básicas en capacidad de almacenamiento y velocidad.
- **Flash memory:** útil para almacenar los avances del proyecto además de documentos, de apoyo al desarrollo.
- **Impresora:** para la impresión de la documentación en cada avance presentado y corrección del mismo.

#### **3.1.1.2.1.3 Recursos Software:**

Los recursos Software que se utilizaron en la elaboración y desarrollo del presente proyecto son:

- **PgAdmin III:** es un gestor de Base de Datos que permite la administración, mediante una interfaz de usuario muy amigable, además de poseer varios formatos para importar como para exportar datos, permite la ejecución de BD robustas o con gran número de registros
- **Rapid Miner:** herramienta que permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico, está bajo la licencia GNU GPL, que permite el acceso gratuito y sin restricciones a las características de la misma, es multiplataforma y proporciona más de 500 operadores orientados al análisis de datos.

#### **3.1.1.2.1.4 Datos:**

- La recopilación de datos provino de 2 formas distintas la primera en cuanto a la EDD se lo hizo de forma física a través de un Cd, mientras que las calificaciones de estudiantes se los obtuvo por medio del web services de la Universidad.

### 3.1.1.2.1.5 Materiales:

- Los materiales de oficina fueron muy útiles en el presente proyecto, ya que ayudaron a la elaboración de documentos en los avances, supervisión y corrección de los mismos el material necesario fueron hojas, esferos, borradores, carpetas, etc.

### 3.1.1.2.2 Riesgos y contingencias.

Los riesgos dentro de un proyecto existen, es por eso que dentro de cualquier planificación se debe prever los mismos y elaborar el o los planes de contingencia necesarios, para evitar contratiempos futuros en el desarrollo.

La tabla XXXIX muestra una lista de los posibles riesgos en la ejecución del proyecto y los planes de contingencia oportunos para cada situación mencionada.

Tanto los riesgos mencionados como los planes de contingencia tuvieron por finalidad reducir el impacto dentro del proyecto, si llega a ocurrir determinado evento o riesgo.

TABLA XXXIX: RIESGOS Y CONTINGENCIAS DEL PROYECTO.

Riesgos	Contingencias
Datos incompletos	Realizar la verificación de los datos obtenidos, mediante la inspección y observación de los mismos.
Inconvenientes en el software o aplicaciones utilizadas	Para prevenir este fallo es necesario realizar, ejecuciones constantes, de tal manera que se pueda sondear, la capacidad y funcionamiento de los mismos.
Perdida o daño de información	Hacer los respaldos de información continuamente en diferentes dispositivos de almacenamiento.
Retrasos y postergación de actividades en la planificación de los tiempos	Tomar en cuenta cada aspecto del desarrollo que pueda causar un contratiempo y a la vez estimar un

	tiempo de solución del mismo.
Acumulación de trabajo pendiente en etapas de desarrollo anteriores	Prever el tiempo y recursos necesarios para la ejecución de cada actividad dentro de una fase, analizando detenidamente la culminación exitosa de las mismas.
Inconvenientes en los equipos	Realizar mantenimiento constante a los equipos de tal manera que se pueda llevar a cabo con normalidad cada uno de los procesos.
Generación de modelos inválidos	Analizar minuciosamente la selección de los atributos para que posteriormente no haya sesgo de información y por ende generación de modelos de MD inválidos.
Pérdida de tiempo en tareas irrelevantes	Determinar con claridad lo que se desea alcanzar y centrar la concertación en la ejecución de lo importante.
Procesos inconclusos	Verificar que todo el desarrollo tenga una secuencia lógica, para que no queden ideas y procedimientos inconclusos.

### 3.1.1.2.3 Terminología

La terminología ayudó de una mejor manera a la comprensión de la información existente.

#### Términos del Negocio

- **EDD:** Evaluación de Desempeño Docente
- **UTI:** Unidad de Telecomunicaciones e Información

- **Web Services:** es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet [101].
- **Batch:** es un archivo de procesamiento por lotes. Se trata de archivos de texto sin formato, guardados con la extensión BAT que contienen un conjunto de comandos lotes [102].
- **GNU/GPL:** Licencia Pública General de más ampliamente usada en el mundo del software y garantiza a los usuarios la libertad de usar, estudiar, compartir (copiar) y modificar el software [103].

### **Términos de Minería de Datos**

- **MD:** Minería de Datos es la ciencia de extracción de información útil y descubrimiento de conocimiento en grandes conjuntos de datos o de bases de datos [104].
- **SGBD:** Sistemas de Gestión de Base de Datos, es una colección de datos relacionados entre sí, estructurados y organizados, y un conjunto de programas que acceden y gestionan esos datos [105].
- **Técnicas de Minería de Datos:** Las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados [106].
- **KDD:** Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases), es el proceso no trivial de identificar patrones válidos, nuevos, potencialmente útiles y en última instancia comprensible en los datos [107].
- **Conocimiento:** Conjunto de experiencias, saberes, valores, información, percepciones que crean e incorporan nuevas ideas, saber y experiencias. [108].

### 3.1.1.2.4 Presupuesto

El presupuesto se detalla en costes del talento humano, recursos materiales, hardware o equipos, licencias, aplicaciones o software.

#### 3.1.1.2.4.1 Costes de Talento Humano

Son valores en pago de honorarios que corresponden al trabajo realizado, por cada persona que interviene en el desarrollo del proyecto (Ver Tabla XL), estos valores se dan en base al rol de los integrantes del proyecto tomando en cuenta el costo por hora de todos los integrantes del equipo de desarrollo, durante el periodo de desarrollo del proyecto. Cada integrante está definido por un rol que desempeño en el proyecto y las horas de trabajo.

TABLA XL: COSTO POR HORA DE INTEGRANTES DE PROYECTO.

Integrantes	Sueldo/Hora (\$)
Asesor de Proyecto	\$ 5.00
Investigador	\$ 5.00

A continuación se detallan las actividades llevadas a cabo con un estimado de duración en horas (Ver Tabla XLI):

TABLA XLI: ACTIVIDADES DEL PROYECTO Y DURACIÓN.

Nro.	Actividad	Duración (horas)
1	Recopilación y análisis de Datos	250
2	Analizar las técnicas de minería de datos que se ajustan al contexto del problema	95
3	Tratamiento, pre procesado, limpieza y transformación de los datos.	180
4	Generación de modelos e identificación de patrones	300
5	Evaluación del o los modelos generados.	85
6	Validación del o los modelos de MD.	50
	Total:	960

Las actividades desarrolladas dentro del proyecto implican la participación en cierto grado de cada integrante del proyecto (Ver Tabla XLII).

TABLA XLII: ASIGNACIÓN DE ACTIVIDADES POR INTEGRANTE Y NÚMERO DE HORAS.

Actividad	Integrante	% de Participación	Horas
1	Asesor de Proyecto	10 %	25
	Investigador	90 %	225
2	Asesor de Proyecto	10 %	9.5
	Investigador	90 %	85.5
3	Asesor de Proyecto	10 %	18
	Investigador	90 %	162
4	Asesor de Proyecto	25 %	75
	Investigador	75 %	225
5	Asesor de Proyecto	25%	21.25
	Investigador	75%	63.75
6	Asesor de Proyecto	25%	12.5
	Investigador	75%	37.5

A continuación, se sumaron las horas de cada uno de los integrantes que intervinieron en el desarrollo del proyecto de esta forma poder calcular el coste del personal por hora invertidas (Ver Tabla XLIII).

TABLA XLIII: RECOPIACIÓN DE HORAS Y COSTES POR ROL DEL PERSONAL

Integrante	Sueldo por hora (\$)	Horas invertidas	Coste Total (\$)
Asesor de Proyecto	5.00	1400	\$ 7000.00
Investigador	5.00	280	\$ 1400.00
Total:			\$ 8400

### 3.1.1.2.4.2 Costes Recursos Hardware

Los equipos utilizados para el desarrollo del presente proyecto investigativo fueron los detallados a continuación con su respectivo coste y depreciación.

La depreciación de los equipos o herramientas es de un 10% anual, tomando en cuenta el tiempo de vida útil que son 4 años, se hace una diferencia de uso desde su adquisición (Ver Tabla XLIV).



TABLA XLIV: COSTES DEL HARDWARE.

Hardware	Precio Unitario (\$)	Depreciación= PU/0.1*4	Precio Total (\$)
<b>1 Computador Personal.</b>	\$ 1130,00	452	\$ 678
<b>1 Impresora</b>	\$ 56,00	22,4	\$ 33.6
<b>2 flash memory</b>	\$ 12.00	4.8	\$ 7.2
			\$ 718,8

Descripción de los equipos utilizados:

- **Computador Personal:** Intel Core 2 Quad 2.4 Ghz, Disco Duro Samsung 1TB, DVD-RW, MBO INTEL De31 Pr s775, Modem: Fax 56 kbps, Sistema Operativo Windows 7; 32 bits, Monitor: Le LCD 20”.
- **Impresora:** HP multifunción, conectividad USB 2.0.
- **Flash memory:** Kingston 4 GB

#### 3.1.1.2.4.3 Costes Recursos Software

Los costos en cuanto al software dependen del tipo de licencia asociada al uso de los mismos, se detallan a continuación (Ver Tabla XLV).

TABLA XLV: COSTES DEL SOFTWARE.

Software	Precio Unitario (\$)	Coste Total (\$)
<b>Rapid Miner</b>	\$ 0.00	\$ 0.00
<b>Pg. Admin III</b>	\$ 0.00	\$ 0.00
		\$ 0.00

El software utilizado es de uso libre por lo que no incurrió en gastos dentro del desarrollo del proyecto.

#### 3.1.1.2.4.4 Materiales y Servicios

Los materiales y servicios fueron indispensables y fundamentales para la elaboración del proyecto, los materiales debido a que ayudaron a la documentación y preparación de los avances y los servicios porque ayudaron en la ejecución y espacio de trabajo del investigador (Ver Tabla XLVI).

Los materiales de oficina como papel, tinta esferos, grapadora, calculadora etc., proporcionaron un trabajo organizado, al momento de realizar la entrega de avances del proyecto. En cambio los servicios básicos como el suministro de luz, teléfono, internet o transporte facilitaron la ejecución y supervisión del mismo.

Se toman en cuenta siguientes gastos:

- Abastecimiento eléctrico los diez meses que duró el desarrollo del proyecto.
- Acceso permanente a internet
- Transporte para revisiones periódicas del desarrollo del proyecto.

TABLA XLVI: COSTES MATERIALES Y SERVICIOS

Materiales y Servicios	Descripción	Costo por Mes	Costo Desarrollo del Proyecto
Materiales de oficina	Papel, esferos, carpetas, cartuchos, etc.	20,00	200,00
Servicios	Luz	15,00	150,00
	Acceso a internet	20,00	200,00
	teléfono	10,00	100,00
Total			650,00

Sumando los costes de desarrollo del proyecto se obtuvo un resultado de \$ 650,00. Se debe tomar en cuenta que además de los materiales y servicios, también se consideró más gastos que incurren la ejecución del proyecto, como transporte, capacitación, entre otros (Ver Tabla XLVII).

TABLA XLVII: GASTOS ADICIONALES.

Gastos	Descripción	Coste Total (\$)
<b>Transporte</b>	Movilización	\$ 30,00
<b>Publicación de Resultados.</b>	Publicación en revista indexada.	\$ 00,00
Total:		\$ 30,00

### 3.1.1.2.4.5 Coste Total

En esta sección se recopilamos todos los gastos antes descritos, para realizar el cálculo final de los costes de ejecución del proyecto (Ver Tabla XLVIII).

TABLA XLVIII: COSTE TOTAL DEL PROYECTO

Recursos	Coste Total (\$)
<b>Costes de Personal</b>	\$ 8400
<b>Costes de Hardware</b>	\$ 718,8
<b>Costes de Software</b>	\$ 0,00
<b>Costes de Material y Servicios.</b>	\$ 650,00
<b>Gastos adicionales</b>	\$ 30,00
Total:	\$ 9798.8

### 3.1.1.2.5 Cronograma del Proyecto

El cronograma del proyecto permitió realizar las actividades necesarias dentro de fechas de asignación estimadas para llevar un desarrollo ordenado (ver figura 22).

























		Nombre	Duración	Inicio
1		<b>☐FASE 1: Recopilación y Análisis de los Datos proporcionados con respecto al proceso de Evaluación al Desempeño Docente</b>	61d?	<b>15/10/2013</b>
2		Realizar entrevistas con el responsable de la Unidad de Telecomunicaciones e Información (UTI) para solicitar autorización en cuanto al acceso de los datos	10d?	15/10/2013
3		Recopilación de los datos necesarios dentro del proceso de evaluación al desempeño docente, en el formato requerido.	20d?	29/10/2013
4		Explorar y analizar los datos disponibles.	15d?	26/11/2013
5		Identificar la situación actual del proceso de Evaluación de Desempeño Docente	10d?	17/12/2013
6		Redacción y Documentación de la Memoria Final.	1d?	31/12/2013
7		Tutoría para la presentación del avance del Trabajo de Titulación.	1d?	07/01/2014
8		Fin Fase 1	0d?	07/01/2014
9		<b>☐FASE 2: Analizar las técnicas de Minería de Datos(MD) factibles para el proceso de Evaluación al Desempeño Docente</b>	49d?	<b>08/01/2014</b>
10		Búsqueda de información en fuentes académicas, revistas, tesis doctorales artículos científicos sobre las diversas técnicas de MD.	20d?	08/01/2014
11		Investigar y evaluar herramientas útiles para el análisis y aplicación de técnicas de MD.	15d?	06/02/2014
12		Selección de las herramientas y algoritmos para el análisis de evaluación docente.	10d?	27/02/2014
13		Redacción y Documentación de la Memoria Final.	2d?	13/03/2014
14		Tutoría para la presentación del avance del Trabajo de Titulación.	1d?	17/03/2014
15		Fin Fase 2	0d?	17/03/2014
16		<b>☐FASE 3: Proceso de Minería de Datos.</b>	36d?	<b>18/03/2014</b>
17		Procesos y Tratamiento de los datos en la elaboración de un modelo de minería de datos	30d?	18/03/2014
18		Redacción y Documentación de la Memoria Final.	5d?	29/04/2014
19		Tutoría para la presentación del avance del Trabajo de Titulación.	1d?	06/05/2014
20		Fin Fase 3	0d?	06/05/2014
21		<b>☐FASE 4: Evaluación del modelo de generado e interpretación de Resultados</b>	23d?	<b>07/05/2014</b>
22		Evaluación Modelo	19d?	07/05/2014
23		Validación del Modelo y Análisis de Resultados	10d?	19/05/2014
24		Redacción y Documentación de la Memoria Final.	4d?	02/06/2014
25		Tutoría para la presentación del avance del PFC.	1d?	06/06/2014
26		Fin Fase 4	0d?	06/06/2014
27		<b>☐FASE 5: Culminación de la redacción de la memoria final del Proyecto Fin de Carrera.</b>	9d?	<b>09/06/2014</b>
28		Redacción de un artículo científico basado en el estándar IEEE.	7d?	09/06/2014
29		Tutoría para la presentación del avance del PFC.	1d?	18/06/2014
30		Culminación del Proyecto Fin de Carrera.	1d?	19/06/2014
31		Fin Fase 5	0d?	19/06/2014

Figura 22: Cronograma de Proyecto.

### **3.1.1.3 Determinación de metas de la minería de datos**

La minería de Datos se define como el proceso de extracción de conocimiento válido, útil, comprensible y desconocido a partir de datos almacenados [20].

Prácticamente hoy en día, no existe una faceta de la realidad de la cual no se disponga de información electrónica, ya sea en bases de datos o en forma textual. Desgraciadamente, gran parte de esa información se construye para un fin concreto pero no se analiza ni se integra con el resto de la información; un ejemplo de esto, son las instituciones que cuentan con una base de datos para el funcionamiento de las aplicaciones del día a día, pero raramente se utiliza con fines analíticos. Esto se debe a que no se sabe cómo hacerlo o no se cuenta con el personal capacitado y las herramientas indicadas para ello.

Cabe indicar que la información reduce nuestra incertidumbre sobre aspectos de la realidad, y por tanto nos permite tomar mejores decisiones, es por eso que actualmente, las grandes empresas e instituciones cuentan con una gran cantidad de datos históricos que detallan fielmente, la vida de esa empresa o institución a lo largo de los años y es mediante la Minería de Datos, que se puede extraer información útil y novedosa que sirva para el perfeccionamiento, supervivencia o mejora de dichas instituciones.

Por lo tanto las metas de la minería de datos radican en la toma de decisiones en base a la información disponible, unas de las metas son:

- Automatización de procesos para la toma de decisiones.
- Generación de estrategias, e incremento de ganancias.
- Minorar las pérdidas en empresas o instituciones públicas o privadas.
- Descubrimiento de comportamientos o patrones que, en primera instancia no son fáciles de detectar.

La meta de la minería de datos en el presente proyecto es que en base a la información disponible como: datos de evaluación de desempeño, notas de estudiantes y técnicas establecidas, se pueda descubrir patrones de comportamiento y factores de interés en cuanto al proceso de EDD, y de esta manera se pueda aportar a la toma de decisiones académicas como administrativas.

#### **3.1.1.4 Plan del Proyecto**

El plan del proyecto ayudó a llevar de forma ordenada el alcance de cada objetivo, pero además permitió detallar cada actividad llevada a cabo con su duración en horas, las dependencias, el porcentaje de participación de cada integrante pero además las herramientas y técnicas utilizadas para la obtención de salidas o resultados de la investigación, el plan se detalla a continuación (Ver Tabla LI)

TABLA XLIX: PLAN DEL PROYECTO.

Actividad	Duración (horas)	Dependencias	Integrantes		Herramientas	Técnicas	Entradas	Salidas
			A	I				
<b>1. Recopilación y Análisis de datos proporcionados por la UTI</b>	130	0	10	90	• Pg Admin III	Selección y Generación de Consultas sql	Datos proporcionados	Bases de datos centralizadas en SGBD y preparadas para ser gestionadas
<b>2. Administrar y gestionar las Bases de Datos</b>	170	1	10	90	• Pg Admin III • Hoja de Cálculo Procesador de Texto	CRISP-DM	Base de Datos Centralizada.	Comprensión, análisis e identificación de la estructura de datos.
<b>3. Analizar técnicas de MD en el desarrollo del proyecto.</b>	70	2	10	70	• Procesador de Texto	CRISP-DM	Fuentes de Información, confiables	Elección de las Técnicas que se ajustan al contexto del problema.
<b>4. Generar modelos</b>	200	3	25	75	• Rapid Miner • Procesador de Texto	Clasificación y Selección de Atributos	Datos procesados, transformados	Elaboración de Modelos en el proceso de EDD
<b>5. Evaluar modelo generado.</b>	155	4	25	75	• Rapid Miner • Procesador de Texto	CRISP-DM	Modelo generado	Redacción de las características de y conocimiento obtenido en el modelo.
<b>6. Validar el o los Modelos generados y análisis de resultados.</b>	155	5	25	75	• Procesador de Texto	CRISP-DM	Modelo evaluado	Validación e Interpretación de resultados, documentación final del proyecto

### **3.1.2 Comprensión de los Datos**

La comprensión de los datos abarca el entendimiento total de la estructura, creación de funciones, vistas y consultas que permitan a su vez el análisis de la información y registros que contiene la BD disponible.

#### **3.1.2.1 Recolección Inicial de los Datos**

Para realizar el proceso de minería de datos, fue necesario que previamente se decida y seleccione los datos que intervendrían en el tratamiento de la información. Los datos tomados pertenecen principalmente a la EDD, como asignaturas dictadas por carreras, estudiantes y docentes por periodo de evaluación, criterios y notas de estudiantes.

La información fue recopilada de 2 fuentes distintas, el proceso de EDD se los obtuvo en físico y las notas de estudiantes, se obtuvo a través del web services de la Universidad. Los datos tomados de diversas fuentes complicaron en cierto grado el trabajo y análisis de los mismos, ya que se requirió unificar todo dentro de una misma base de datos, dando como consecuencia una extensión del trabajo realizado.

La información con la cual se trabajó pertenece al Sistema de Evaluación de Desempeño Docente y al Sistema de Gestión Académica de la Universidad Nacional de Loja, proporcionada por la Unidad de Telecomunicaciones e Información, datos en físico y a través del Web Services. Cabe señalar que los datos, tomados para el presente proyecto pertenecen a los estudiantes y docentes que evaluaron dentro de los periodos de estudio, en el Área de la Energía de las Industrias y los Recursos Naturales No Renovables, que cuenta con un número de 1910 estudiantes en los 3 periodos de evaluación, pertenecientes a las carreras de : Ingeniería en Electromecánica, Ingeniería en Electrónica y Telecomunicaciones, Ingeniería en Geología Ambiental y Ordenamiento Territorial, Ingeniería en Geología Ambiental y Ordenamiento Territorial(especial), Tecnología en Electricidad y Control industrial e Ingeniería en Sistemas, la última con mayor número de estudiantes.

La recopilación de los datos se realizó de distintas formas; en cuanto a la EDD, y las notas de los estudiantes.



Los datos que pertenecen al proceso de EDD, se los obtuvo en físico (Ver Figura 23) por lo tanto para el entendimiento y comprensión de los mismos se hizo una análisis minucioso de la BD proporcionada.

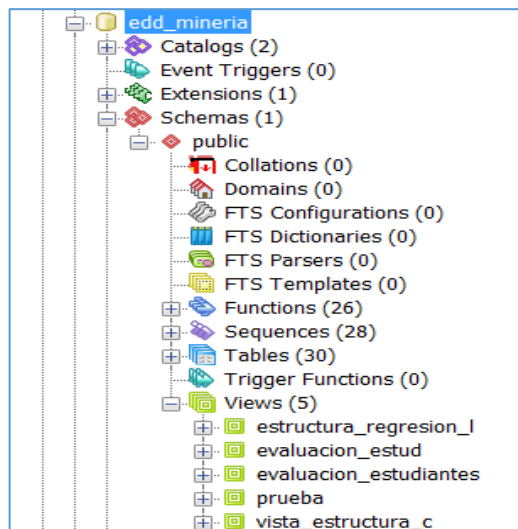


Figura 23: BD en el proceso de EDD.

La BD de análisis se encuentra dentro de un formato administrado por Postgres, y cuenta con 30 tablas con la creación de estructuras útiles en la minería, 5 vistas y la creación de funciones en el análisis y recopilación de registros.

Cabe señalar que los datos correspondientes a periodos de evaluación anteriores se encuentran en un formato de mysql con diferente estructura a la actual, y que los datos que se analizaron dentro del presente trabajo de titulación pertenecen a los periodos 2011, 2012, 2013 del AEIRNNR.

Las notas de los estudiantes, se las obtuvo mediante el web services de la universidad, en donde se puede hacer uso de las siguientes categorías (Ver Tabla L):

TABLA L: DESCRIPCIÓN DE LOS SERVICIOS POR CATEGORÍAS.

<b>Categoría</b>	<b>Descripción</b>
Académica	Esta categoría tiene métodos que devuelven información acerca de estudiantes y docentes.
Institucional	Esta categoría retorna información institucional, como datos de módulos, paralelos, carreras y áreas
Personal	Esta categoría retorna a través de sus métodos, información

	personal de estudiantes y docentes.
Validación	Esta categoría posee métodos que sirven para la validación de docentes y estudiantes dentro de la BD.
Estadística	Esta categoría se retorna información estadística, como número de estudiantes aprobados y número de estudiantes reprobados, entre otros.

Las notas de los estudiantes se las obtuvo, de la categoría Académica. Para obtener las notas de estudiantes se desarrolló una aplicación que ayude a consumir los datos, en el web services. Además la herramienta Pg Admin III sirvió en la gestión de Base de Datos.

### 3.1.2.2 Descripción de los datos.

Los datos adquiridos como se mencionó anteriormente fueron recopilados de dos fuentes de información, como el Sistema de Evaluación de Desempeño Docente y del Sistema de Gestión Académico de la Universidad Nacional de Loja, posteriormente se procedió a limpiar e integrar todos los datos dentro de una Base de Datos que permita llevar a cabo procesos de selección, consultas, y procesos necesarios para la ejecución del presente proyecto. Los datos adquiridos corresponden a periodos académicos desde el 2011 hasta el 2013 en el proceso de EDD.

Para obtener los datos se emplearon diversos, métodos y procesos que permitieron la obtención de tablas descritas a continuación (Ver Tabla LI, LII, LIII, LIV):

- **BASE DE DATOS EDD**

TABLA LI: TABLA CARRERAS Y ASIGNATURAS

Tabla	Descripción
app_areasga	Retorna el nombre y número de áreas en la Universidad Nacional de Loja
app_carrera	Retorna cada carrera de la energía
app_asignatura	Retorna cada una de las asignaturas dictadas
app_asignaturadocente	Retorna las asignaturas por docente
app_estudianteasignaturadocente	Retorna los estudiantes que corresponden a una asignatura dictada por un docente

**TABLA LII: TABLAS ESTUDIANTES Y DOCENTES**

Tabla	Descripción
app_estudianteperiodoacademico	Retorna los estudiantes que corresponden a un determinado periodo académico
estudiante_evaluacion	Retorna los estudiantes que corresponden a un periodo de evaluación determinado
app_docenteperiodoacademico	Retorna los docentes que corresponden a un determinado periodo académico
docente_evaluacion	Retorna los docentes que corresponden a un periodo de evaluación determinado

**TABLA LIII: TABLAS CUESTIONARIO Y EVALUACIÓN.**

<b>Tabla</b>	<b>Descripción</b>
app_cuestionario	Retorna los cuestionarios en los periodos de evaluación
app_seccion	Retorna las secciones que contiene un cuestionario
app_pregunta	Retorna las preguntas dentro de una sección en un cuestionario
app_tipopregunta	Retorna el tipo de pregunta
app_itempregunta	Retorna los ítem en una pregunta
app_contestacion	Retorna la contestación con respecto a una pregunta dentro de una evaluación
app_evaluacion	Devuelve el id de evaluación, con fecha de inicio y fin
app_periodelaevaluacion	Retorna los periodos de evaluación existentes dentro de un periodo académico
app_periodelaevaluacion_areasSGA	Retorna los periodos de evaluación de cada área
app_periodelaacademico	Retorna los periodos académicos en los que se llevan a cabo las evaluaciones.
app_periodelaacademico_ofertasacademicassga	Retorna las ofertas académicas en un periodo académico.
app_ofertaacademicassga	Retorna las ofertas académicas propuestas dentro de un área.
app_tipoinformante	Retorna estudiantes, docentes, directivo o par académico categorizado en el proceso.

app_usuario	Retorna el tipo de usuario que ingresa al sistema
auth_user	Retorna información como nombres, apellidos, email de los usuarios que ingresan al sistema de evaluación.

- **Notas Estudiantes**

De la categoría académica del web services de la universidad, se obtuvo las notas de los estudiantes usando los siguientes métodos:

TABLA LIV: MÉTODOS UTILIZADOS PARA OBTENER INFORMACIÓN DE CATEGORÍA ACADÉMICA.

Método	Descripción
sga_periodos_lectivos()	Este método retorna una lista de registros con todos los periodos lectivos
sgaws_notas_estudiante()	Retorna las notas de un estudiante en cada unidad, la carrera, módulo, paralelo al que pertenece y estado de la matrícula en la oferta académica

La base de datos final contiene las siguientes tablas:

- app\_areasga: contiene el nombre y número de áreas en la Universidad Nacional de Loja
- app\_carrera: contiene las carreras del área de la energía
- app\_asignatura: contiene cada una de las asignaturas dictadas
- app\_asignaturadocente: contiene las asignaturas por docente
- app\_estudiante\_asig\_doc: contiene los estudiantes que corresponden a una asignatura dictada por un docente
- app\_estudianteperiodoacademico: contiene los estudiantes que corresponden a un determinado periodo académico
- estudiante\_evaluacion: contiene los estudiantes que corresponden a un periodo de evaluación determinado

- app\_docenteperiodoacademico: contiene los docentes que corresponden a un determinado periodo académico
- docente\_evaluacion: contiene los docentes que corresponden a un periodo de evaluación determinado
- app\_cuestionario: contiene los cuestionarios en los periodos de evaluación
- app\_seccion: contiene las secciones que contiene un cuestionario
- app\_pregunta: contiene las preguntas dentro de una sección en un cuestionario
- app\_tipopregunta: contiene el tipo de preguntas, de ensayo o selección única
- app\_itempregunta: contiene orden de ítem de preguntas
- app\_contestacion: contiene la contestación con respecto a una pregunta dentro de una evaluación
- app\_evaluacion: contiene el id de evaluación, con fecha de inicio y fin
- app\_periodoevaluacion: contiene los periodos de evaluación existentes dentro de un periodo académico.
- app\_periodoevaluacion\_areasSGA: contiene los periodos de evaluación de cada área
- app\_periodoacademico: contiene los periodos académicos en los que se llevan a cabo las evaluaciones
- app\_periodoacademico\_ofertasacademicassga: contiene las ofertas académicas en un periodo académico.
- app\_ofertaacademicasga: contiene las ofertas académicas propuestas dentro de un área.
- app\_tipoinformante: contiene estudiantes, docentes, directivo o par académico categorizado en el proceso.
- app\_usuario: contiene el tipo de usuario que ingresa al sistema
- auth\_user: contiene nombres, apellidos, email de los usuarios que ingresan al sistema de evaluación.

- nota\_unidad\_estudiante: contiene notas académicas que se han obtenido en todas las materias cursadas por los estudiantes.

A continuación se describe la estructura de cada tabla que se encuentra en la base de datos obtenida (Ver Tabla LV- LXXIX)

TABLA LV: ESTRUCTURA DE TABLA APP\_AREASGA

Atributos	Descripción	Tipo	Llaves
id	Identificador de áreas	serial	PK
siglas	Siglas de cada área	character varying(10)	
nombre	Nombre de cada área	character varying(256)	

TABLA LVI: ESTRUCTURA DE TABLA CARRERA.

Atributos	Descripción	Tipo	Llaves
id	Identificador director	serial	PK
carrera	Nombre de carreras	character varying(255)	

TABLA LVII: ESTRUCTURA DE TABLA APP\_ASIGNATURA.

Atributos	Descripción	Tipo	Llaves
id	Identificador asignaturas	serial	PK
area	Siglas de cada área	character varying(20)	
carrera	Nombre de cada área	character varying(100)	
semestre	Número de Módulos	character varying(10)	
paralelo	Nombre paralelo	character varying(50)	
nombre	Nombre de la asignatura	text	
Periodo_academico_id	Id periodo académico	integer	FK-> app_periodoacademico

TABLA LVIII: ESTRUCTURA DE TABLA APP\_ASIGNATURADOCENTE

Atributos	Descripción	Tipo	Llaves
id	Identificador asignatura docente	serial	PK

asignatura_id	Identificador de asignaturas	integer	FK-> app_asignatura
docente_id	Identificador de docente	integer	FK-> app_docenteperiodoacademico

TABLA LIX: ESTRUCTURA DE TABLA APP\_ESTUDIANTE\_ASIG\_DOC

Atributos	Descripción	Tipo	Llaves
id	Identificador estudiante asignatura docente	serial	PK
estudiante_id	Identificador estudiantes	integer	FK-> app_estudianteperiodoacademico
asig_doc_id	Identificador asignaturas docente	integer	FK-> app_asignaturadocente
nota	Calificación de asignatura		

TABLA LX: ESTRUCTURA DE TABLA APP\_ESTUDIANTEPERODOACADEMICO

Atributos	Descripción	Tipo	Llaves
id	Identificador estudiante periodo académico	serial	PK
usuario_id	Identificador usuarios	integer	FK-> app_usuario
periodo_academico_id	Identificador periodo académico	integer	FK-> app_periodoacademico

TABLA LXI: ESTRUCTURA DE TABLA ESTUDIANTE\_EVALUACION

Atributos	Descripción	Tipo	Llaves
cedula	Cedula estudiante	character(15)	
id_est_period_acad	Id estudiante p.a	integer	PK FK-> app_estudianteperiodoacademico
nombres	Nombres estudiantes	character varying(50)	
apellidos	Apellidos estudiantes	character varying(50)	
email	Correo electrónico	varchar(50)	
clave	Clave	varchar(50)	

TABLA LXII: ESTRUCTURA DE TABLA APP\_DOCENTE PERIODO ACADEMICO

Atributos	Descripción	Tipo	Llaves
id	Identificador docente periodo académico	serial	PK
usuario_id	Identificador usuarios	integer	FK-> app_usuario
periodo_academico_id	Identificador periodo académico	integer	FK-> app_periodoacademico
par_academico	Condición: verdadero-falso	boolean	

TABLA LXIII: ESTRUCTURA DE TABLA DOCENTE\_EVALUACION

Atributos	Descripción	Tipo	Llaves
id_doc_period_acad	Identificador docente periodo académico	integer	PK
cedula	Cédula docente	character varying(15)	
nombres	Nombres docente	character varying(100)	
apellidos	Apellidos docente	character varying(100)	
email	varchar(100)		
clave	varchar(100)		
titulo	varchar(100)		

TABLA LXIV: ESTRUCTURA DE TABLA APP\_CUESTIONARIO

Atributos	Descripción	Tipo	Llaves
id	Identificador cuestionario	serial	PK
titulo	Titulo cuestionario	character varying(255)	
informante_id	Identificador, tipo de informante	integer	FK-> app_tipoinformante
periodo_evaluacion_id	Identificador periodo de evaluación	integer	FK-> app_periodoevaluacion
nombre	Nombre de cuestionario	character varying(150)	
peso	Valor de cuestionario	double	



TABLA LXV: ESTRUCTURA DE TABLA APP\_SECCION

<b>Atributos</b>	<b>Descripción</b>	<b>Tipo</b>	<b>Llaves</b>
id	Identificador sección	serial	PK
titulo	Título sección	character varying(200)	
descripcion	Descripción de sección	text	
orden	Orden de secciones	integer	
cuestionario_id	Identificador cuestionario	integer	FK-> app_cuestionario
codigo	Código secciones	character varying(20)	
superseccion	Sección que contiene más secciones		
ponderación	Peso por tipo de informantes		

TABLA LXVI: ESTRUCTURA DE TABLA APP\_PREGUNTA

<b>Atributos</b>	<b>Descripción</b>	<b>Tipo</b>	<b>Llaves</b>
id	Identificador pregunta	serial	PK
texto	Texto pregunta	text	
orden	Orden de preguntas	integer	
tipo_id	Identificador tipo de pregunta	integer	FK-> app_tipopregunta
seccion_id	Identificador sección	integer	FK-> app_seccion

TABLA LXVII: ESTRUCTURA DE TABLA APP\_TIPOPREGUNTA

<b>Atributos</b>	<b>Descripción</b>	<b>Tipo</b>	<b>Llaves</b>
id	Identificador	integer	PK
tipo	Tipo de pregunta	varchar(20)	
descripcion	Descripcion de tipo	varchar(100)	

TABLA LXVIII: ESTRUCTURA DE TABLA APP\_ITEMPREGUNTA

<b>Atributos</b>	<b>Descripción</b>	<b>Tipo</b>	<b>Llaves</b>
id	Identificador	integer	PK
texto	Escala de calificación	varchar(50)	
pregunta_id	Id pregunta	integer	FK->app_pregunta
orden	Orden de ítem	integer	

descripcion	Descripción de campo	varchar(70)	
-------------	----------------------	-------------	--

**TABLA LXIX: ESTRUCTURA DE TABLA APP\_CONTESTACION**

Atributos	Descripción	Tipo	Llaves
id	Identificador contestación	serial	PK
pregunta	Numero pregunta	integer	
respuesta	Ítem de calificación	text	
evaluacion_id	Identificador, evaluación	integer	FK-> app_evaluacion

**TABLA LXX: ESTRUCTURA DE TABLA APP\_EVALUACION**

Atributos	Descripción	Tipo	Llaves
id	Identificador evaluación	serial	PK
cuestionario_id	Identificador cuestionario	integer	FK-> app_cuestionario
est_asig_docente_id	Identificador asignatura docente	integer	FK-> app_estudianteasignaturadocente
doc_period_acad_id	Identificador docente periodo académico	integer	FK-> app_docenteperiodoacademico
fecha_inicio	Fecha inicio evaluación		
fecha_fin	Fecha inicio evaluación		
hora_inicio	Tiempo de inicio evaluación		
hora_fin	Tiempo fin evaluación		

**TABLA LXXI: ESTRUCTURA DE TABLA APP\_PERIODOEVALUACION**

Atributos	Descripción	Tipo	Llaves
id	Identificador periodo de evaluación	serial	PK
nombre	Nombre periodo evaluación	Character varying(100)	
descripcion	Descripción periodo evaluación	text	
periodo_academico_id	Identificador	integer	FK->

	periodo académico		app_periodoacademico
--	-------------------	--	----------------------

TABLA LXXII: ESTRUCTURA DE TABLA APP\_PERIODOEVALUACION\_AREASSGA

Atributos	Descripción	Tipo	Llaves
<b>id</b>	Identificador	integer	PK
<b>periodoevaluacion_id</b>	Id periodo de evaluación	integer	FK->app_periodoevaluacion
<b>areasga_id</b>	Id área	integer	FK->app_areasga

TABLA LXXIII: ESTRUCTURA DE TABLA APP\_PERIODOACADEMICO

Atributos	Descripción	Tipo	Llaves
id	Identificador periodo académico	serial	PK
nombre	Nombre periodo académico	Character varying(50)	
periodo_lectivo	Año periodo lectivo	character varying(100)	

TABLA LXXIV: ESTRUCTURA DE TABLA APP\_PERIODOACADEMICO\_OFERTASACADEMICASSGA

Atributos	Descripción	Tipo	Llaves
<b>id</b>	Identificador	integer	PK
<b>periodoacademico_id</b>	Id periodo académico	integer	FK->app_periodoacademico
<b>ofertaacademicasga_id</b>	Id oferta académica	integer	FK->app_ofertaacademicasga

TABLA LXXV: ESTRUCTURA DE TABLA APP\_OFERTAACADEMICASGA

Atributos	Descripción	Tipo	Llaves
id	Identificador oferta académica	serial	PK
descripcion	Descripción oferta académica	Character varying(100)	

TABLA LXXVI: ESTRUCTURA DE TABLA APP\_TIPOINFORMANTE

Atributos	Descripción	Tipo	Llaves
id	Identificador informante tipo	serial	PK
tipo	Tipo de informante	Character varying(50)	
descripcion	Descripción tipo de informante	character varying(200)	

TABLA LXXVII: ESTRUCTURA DE TABLA APP\_USUARIO

Atributos	Descripción	Tipo	Llaves
user_ptr_id	Identificador informante	serial	PK FK-> auth_user
cedula	Tipo de informante	Character varying(15)	
titulo	Descripción tipo de informante	character varying(100)	

TABLA LXXVIII: ESTRUCTURA DE TABLA AUTH\_USER

Atributos	Descripción	Tipo	Llaves
id	Identificador usuario	serial	PK FK-> auth_user
username	Cédula usuario	Character varying(30)	
first_name	nombres usuario	Character varying(50)	
last_name	apellidos usuario	character varying(50)	

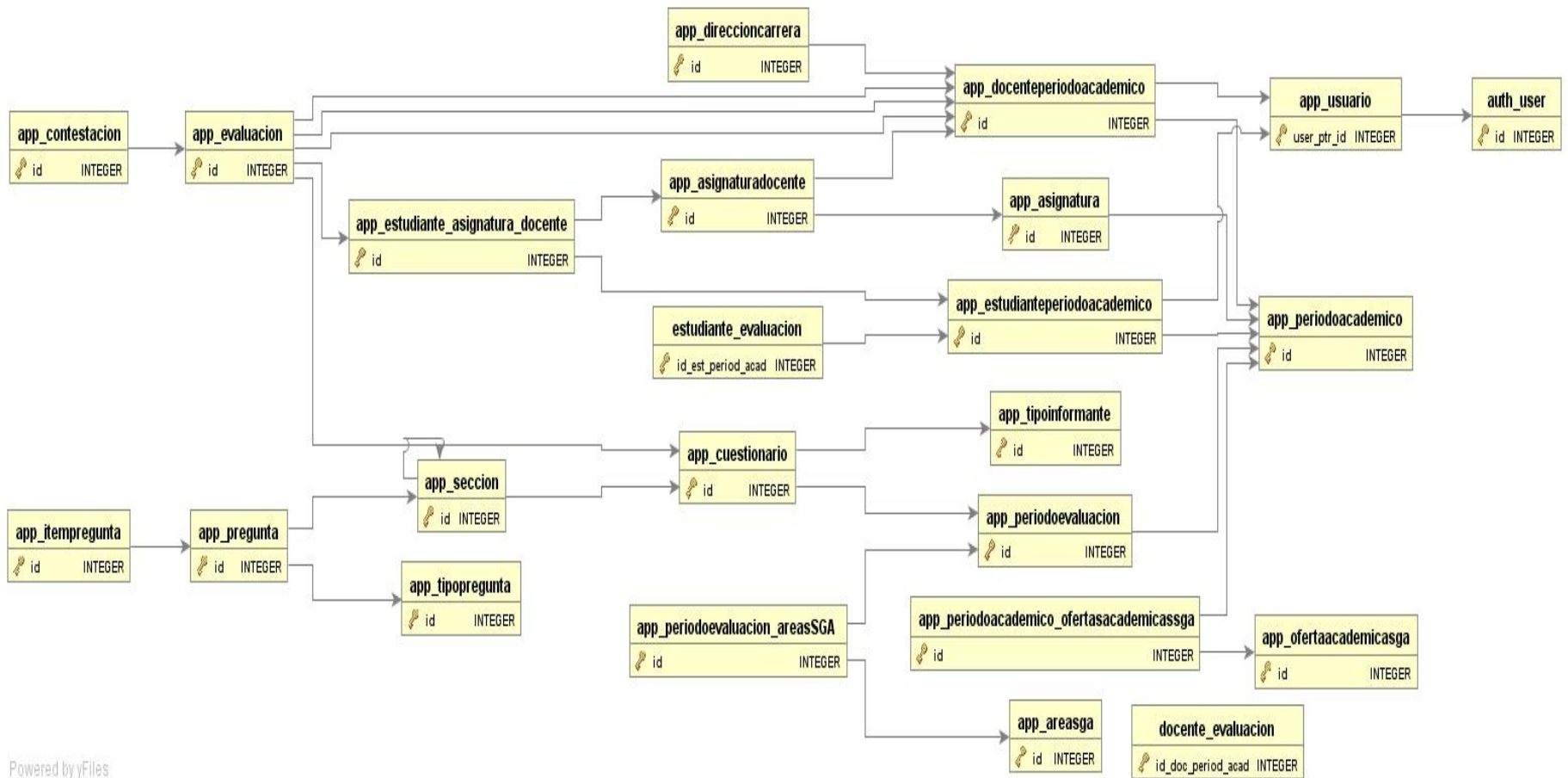
TABLA LXXIX: ESTRUCTURA DE TABLA NOTA\_UNIDAD\_ESTUDIANTE

Atributos	Descripción	Tipo	Llaves
id	Identificador nota	integer	PK
unidad	Asignatura	varchar(250)	
estudiante	Cedula estudiante	varchar(15)	
valor	Calificación	float	
semestre	Numero semestre	char(250)	
paralelo	Nombre paralelo	char(50)	

En base a la información obtenida y al diagrama físico del modelo relacional, de la base de datos analizada. Se pudo determinar las dependencias entre entidades y el tipo de información que contienen.

El diagrama físico del modelo relacional (Ver Figura 24), consiste en mostrar gráficamente las entidades, sus atributos y la relación que existe entre las entidades. Este diagrama corresponde a la base de datos proporcionada por la Unidad de Telecomunicaciones e Información (UTI).

➤ Estructura de la BD analizada



Powered by yFiles

Figura 24: Estructura de la Base de Datos EDD

### 3.1.2.3 Exploración de los datos

En esta etapa se realizó una exploración de los datos y en base a ello un análisis estadístico acerca del proceso de evaluación.

#### **Estudiantes Carreras:**

Respecto al Área de la Energía las Industrias y los recursos Naturales no renovables de la Universidad Nacional de Loja, la carrera que presenta el mayor número de estudiantes en cada evaluación, es la de Ingeniería en Sistemas (Ver Tabla LXXX, Ver figura 24).

TABLA LXXX: NÚMERO DE ESTUDIANTES POR CARRERAS.

<b>Carreras</b>	<b>Septiembre 2011- Julio2012</b>	<b>Septiembre 2012-Febrero 2013</b>	<b>Marzo-Julio 2013</b>	<b>Total Estudiantes</b>
Ingeniería en Electromecánica.	185	120	178	483
Ingeniería en Geología Ambiental y Ordenamiento Territorial	104	87	105	296
Ingeniería en Sistemas	274	223	232	729
Ingeniería en Electrónica y Telecomunicaciones	106	96	123	325
				1833

El resultado es un total de 1833 en los 3 periodos académicos, sumando 77 estudiantes de las carreras especiales dando un resultado final de 1910.

#### **Docentes Carreras:**

De la misma forma que se tomó los estudiantes en los periodos de evaluación, se tomó el total de docentes de cada carrera en los periodos de evaluación (Ver Tabla LXXXIII).

TABLA LXXXI: TOTAL DOCENTES.

Docentes por Periodo de Evaluación				
Carrera	Septiembre 2011 - Julio 2012	Septiembre 2012 - Febrero 2013	Marzo-Julio 2013	TD
Ingeniería en Electromecánica	21	18	21	60
Ing. Geología Ambiental y Ordenamiento Territorial	12	11	14	37
Ingeniería en Sistemas	21	22	19	62
Ingeniería en Electrónica y Telecomunicaciones	12	9	11	32
Total	66	60	65	191

La tabla refleja que en la carrera de ingeniería en sistemas e ingeniería en electromecánica existe la mayor cantidad de docentes, pero por periodo académico tienen una similitud en número.

### Evaluaciones

A continuación se presenta las evaluaciones recopiladas dentro de las bases de datos disponibles por periodo de evaluación (Ver Tabla LXXXII).

TABLA LXXXII: TOTAL EVALUACIONES.

Total de Evaluaciones por Carrera y Periodo Académico									
Carreras	2011-2012			2012-2013			2013-2014		
	E	D	Total	E	D	Total	E	D	Total
Ingeniería en Electromecánica	185	21	3885	120	18	2160	178	21	3738
Ing. Geología Ambiental y Ordenamiento Territorial	104	12	1248	87	11	957	105	14	1470
Ingeniería en Sistemas	274	21	5754	223	22	4906	232	19	4408
In. Electrónica y Telecomunicaciones	106	12	1272	96	9	864	123	11	1353
Total	669	66	<b>12159</b>	526	60	<b>8887</b>	638	65	<b>10969</b>

Sumando los resultados de los 3 periodos académicos da un aproximado de 32015 registros dentro del proceso de evaluación.

La exploración de los datos de esta sección se encuentra en mejor detalle en la Fase 1: Apartado 1.3: Disposición de los datos.

#### **3.1.2.4 Verificación de la calidad de los datos.**

La calidad de los datos es un aspecto determinante del proceso de obtención y validación de un modelo de minería de datos, por lo tanto se verificó la calidad de los mismos con la finalidad de corregir errores, completar campos nulos o en su defecto eliminarlos, esto con el objetivo de realizar el proceso de limpieza.

Aspectos para tomar en cuenta:

- Al recopilar los datos del web services, se dio una incompatibilidad en el formato de los datos con los de la base de datos de EDD.

Posible solución: realizar la migración de los datos, en este caso notas existentes en el web services de la universidad en mysql al formato de datos en postgres.

- Para añadir las notas a la base de datos de EDD, debe haber consistencia en los registros de tal manera que se permita enlazar las tablas de interés, la tabla app\_asignaturas de la base de datos de EDD, contiene la columna id, pero la tabla unidad de la base de datos recopilada del web services carece de identificador dando solo los nombres de cada asignatura, ocurriendo una inconsistencia en los datos

Posible solución: estandarizar los nombres de las materias.

#### **3.1.3 Preparación de los Datos**

La fase de preparación de datos cubrió todas las actividades necesarias para construir el conjunto de datos final (estructuras de minería). Las tareas de preparación de datos pueden ser realizadas muchas veces y en cualquier orden. Las tareas incluyeron la selección de tablas, registros, y atributos, así como la transformación y la limpieza [21].



La fase de preparación de datos quizá es la actividad más importante de las fases descritas, dentro de esta fase se realizó la selección y la limpieza de los datos, armando una estructura adecuada para la explotación de los mismos, de esta estructura dependió el modelo que generó la herramienta y técnicas de minería de datos seleccionadas. Por lo tanto, se realizó este proceso para un buen aprovechamiento de la base de datos.

### **3.1.3.1 Selección de los Datos**

Como se mencionó anteriormente, el análisis radica en la Evaluación de Desempeño Docente del Área de Energía las Industrias y los Recursos Naturales No Renovables, por lo tanto como primer paso, se realizó un filtro de los datos que pertenecen a esta área.

Es decir luego de examinar la base de datos proporcionada por la UTI, se eliminó los registros que no estaban dentro del alcance de estudio, dando aproximado de 1910 estudiantes, 191 docentes y 32.015 evaluaciones en los 3 periodos académicos

Además se optó por eliminar tablas, que no influyeron en la generación del modelo de minería de datos, las tablas anuladas son: app\_periodoevaluacion\_areasSGA, app\_periодоacademico\_ofertasacademicassga, app\_tipopregunta, app\_ofertaacademicasga.

### **3.1.3.2 Limpieza de los Datos**

Luego de la selección de tablas que intervendrán en el proceso de minería, se realizó la limpieza de los datos como eliminación de atributos que contienen datos nulos o tratamiento de valores ausentes, reducción del volumen de datos, eliminar registros duplicados y completar campos vacíos. En esta sección también se presenta la solución a los problemas detectados en la fase dos, tarea cuatro:

- Un problema fue encontrado al recopilar los datos tanto de los web services y la base de datos de EDD, en donde se dio una incompatibilidad en el formato de los datos, la solución realizada fue la migración de los datos existentes en el web services de la universidad en mysql al formato de datos de EDD en postgres.
- Para añadir las notas a la base de datos de EDD, debió haber consistencia en los registros de tal manera que se permita enlazar las tablas de interés, la tabla

app\_asignaturas de la base de datos de EDD, contiene la columna id, pero la tabla unidad de la base de datos recopilada del web services carece de identificador, dando solo los nombres de cada asignatura, ocurriendo una inconsistencia en los datos la solución fue estandarizar los nombres de las materias, reemplazando el código unicode existente por las letras faltantes, y de esta manera agregar las notas de los estudiantes en cada una de las asignaturas a la base de datos final.

- Al realizar la migración de las notas de los estudiantes, a la BD actual de EDD ocurrió que algunas de las asignaturas recibidas por los estudiantes tenían el valor nulo, por lo tanto se filtró la información por estudiante para llenar estos campos de forma manual o en su defecto eliminarlas.

La limpieza básicamente consiste en la eliminación de datos con errores, columnas con registros nulos y la corrección de errores dentro de la base de datos como la duplicidad.

Por ejemplo en la siguiente tabla (Ver Tabla LXXXIII) se describen los datos y registros eliminados:

TABLA LXXXIII: LIMPIEZA DE DATOS

<b>Tabla</b>	<b>Columna</b>	<b>Descripción</b>
app_periodoevaluacion	descripcion	Contiene información irrelevante para el modelado
app_seccion	superseccion	Contiene secciones repetidas
docente_evaluacion	email, clave	Información irrelevante para el modelado
estudiante_evaluacion	email, clave,	Información irrelevante para el modelado

Cabe destacar que el proceso de limpieza es un proceso continuo y retroalimentativo por lo tanto se limpió los datos cuando fue necesario.

### **3.1.3.3 Construcción de Datos**

En lo que concierne a construcción de datos se preparó 4 estructuras (Ver Tabla LXXXV, LXXXVI, LXXXVII), de minería de datos, para cada factor crítico de éxito, considerando los atributos descritos en la siguiente tabla (Ver Tabla LXXXIV), se crearon atributos

derivados con el objetivo de aplicar diferentes modelos de minería de datos, y evaluar el comportamiento de cada modelo.

### Descripción de atributos de Estructuras:

TABLA LXXXIV: ATRIBUTOS EN LA ESTRUCTURA DE MINERÍA DE DATOS.

<b><i>cedula_docent:</i></b>
Atributo de tipo varchar, continuo, y corresponde a los números de identificación de cada docente del AEIRNNR. Esto con el fin de agrupar a los mismos para posteriormente conocer el rendimiento y la evolución de cátedra que tienen.
<b><i>cedula_estudiante</i></b>
Atributo de tipo varchar, continuo, y corresponde a números de identificación de cada estudiante del AEIRNNR.
<b><i>carrera:</i></b>
Atributo de tipo varchar, discreto, útil para generar información acerca del desarrollando que tienen los docentes dentro de determinada carrera, las mismas que pueden ser: <ul style="list-style-type: none"> <li>• Ing. en Elect</li> <li>• Ing en Geol y OT</li> <li>• Ing. en Sist</li> <li>• Ing. Elect y Tele</li> <li>• Ing en Geol y OT_e</li> <li>• Tec_ Elec_CI</li> </ul>
<b><i>asignatura:</i></b>
De tipo varchar, continuo es un atributo muy importante dentro del análisis, motivo por el cual se toman todas las asignaturas impartidas dentro del AEIRNNR, para posteriormente conocer patrones de comportamiento.
<b><i>periodo_acad_id:</i></b>
Atributo de tipo integer, discreto, útil en la clasificación de tiempos correspondientes a cada proceso de evaluación. Puede tomar valores binarios en diferentes estructuras
<b><i>calf_4_3:</i></b>
De tipo integer, atributo que permite conocer los items más altos como respuesta de la evaluación de desempeño docente.
<b><i>calf_2_1:</i></b>
De tipo integer, atributo que permite conocer los items más bajos como respuesta de la evaluación de desempeño docente.
<b><i>cpf_4_3, cpg_4_3, pv_4_3:</i></b>
De tipo integer, atributo que permite conocer los items más altos como respuesta de la evaluación de desempeño docente en cada componente de evaluación.
<b><i>cpf_2_1, cpg_2_1, pv_2_1:</i></b>
De tipo integer, atributo que permite conocer los items más bajos como respuesta de la

evaluación de desempeño docente en cada componente de evaluación.
<b>porcentaje:</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación obtenido por un docente, en una asignatura, calculado de la multiplicación entre <i>calf_4_3</i> por 100 dividido entre el total de items calificados o <i>resultado</i> .
<b>evaluacion:</b>
De tipo varchar, atributo que discretiza el porcentaje de evaluación obtenido por un docente en: Criter-Int: 0-69% Buen-Rend: 70-100%
<b>resul_cpf</b>
De tipo integer, continuo, atributo que muestra el total de <i>cpf_4_3</i> y <i>cpf_2_1</i> , total de capacidad profesional.
<b>resul_cpg</b>
De tipo integer, continuo, atributo que muestra el total de <i>cpg_4_3</i> y <i>cpg_2_1</i> , total de capacidad pedagógica.
<b>resul_pv</b>
De tipo integer, continuo, atributo que muestra el total de <i>pv_4_3</i> y <i>pv_2_1</i> , total de práctica de valores.
<b>porc_cpf</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación en capacidad profesional obtenido por un docente, en una asignatura, calculado de la multiplicación entre <i>cpf_4_3</i> por 100, dividido para <i>resul_cpf</i> . Se discretiza en: regular: 0-69% buena: 70- 80% excelente: 81-100%
<b>porc_cpg</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación en capacidad pedagógica obtenido por un docente, en una asignatura, calculado de la multiplicación entre <i>cpg_4_3</i> por 100, dividido para <i>resul_cpg</i> . Se discretiza en: regular: 0-69% buena: 70- 80% excelente: 81-100%
<b>porc_pv</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación en práctica de valores obtenido por un docente, en una asignatura, calculado de la multiplicación entre <i>pv_4_3</i> por 100, dividido para <i>resul_pv</i> . Se discretiza en: regular: 0-69% buena: 70- 80% excelente: 81-100%
<b>catedra</b>

<p>De tipo varchar, atributo que permite conocer la calificación obtenida por un docente dentro de un periodo académico basado en los ítems más bajos de calificación (2 y1 ) calculado de la multiplicación entre <i>calf_2_1</i> por 100, dividido para el total de ítems calificados o <i>resultado</i>, obteniendo un porcentaje del 0 al 11% y del 12 al 100%. Se discretiza en:</p> <p><b>P:</b>positiva <b>N:</b> negativa</p>
<p><b><i>catedra_cpf, catedra_cpg, catedra_pv</i></b></p>
<p>De tipo varchar, atributo que permite conocer la calificación obtenida por un docente dentro de un periodo académico por cada componente de evaluación puede ser:</p> <p><b>N:</b> negativa <b>P:</b>positiva</p>
<p><b><i>catedra_p_1, catedra_p_2, catedra_p_3</i></b></p>
<p>De tipo varchar, atributo que permite conocer la calificación de un docente por periodo académico basado en el atributo <i>catedra</i>, puede ser:</p> <p><b>N:</b> negativa <b>P:</b>positiva <b>A:</b> <i>ausencia en determinado periodo académico</i></p>
<p><b><i>resultado, total_calf:</i></b></p>
<p>De tipo integer, continuo, atributo que muestra el total de <i>calf_4_3</i> y <i>calf_2_1</i>, total de preguntas en un cuestionario.</p>
<p><b><i>calif_cualitativa:</i></b></p>
<p>De tipo varchar, discreto, atributo que muestra el rendimiento de un docente, basado en los resultados (<i>calf 4 y 3</i>) de las evaluaciones en una materia determinada. El mismo que pueden ser:</p> <ul style="list-style-type: none"> <li>• Muy alto: 81-100%</li> <li>• Alto: 70-80%</li> <li>• Bajo: 66-69%</li> <li>• Muy bajo: 0-65%</li> </ul>
<p><b><i>evol_catedra</i></b></p>
<p>Atributo de tipo discreto, denota la evolución de cátedra que ha tenido un docente en los periodos académicos de estudio, clasificado en:</p> <ul style="list-style-type: none"> <li>• Si</li> <li>• No</li> </ul>
<p><b><i>calif_cualitativa:</i></b></p>
<p>De tipo varchar, discreto, atributo que muestra el rendimiento de un docente, basado en los resultados (<i>item 4 y 3</i>) de las evaluaciones en una materia determinada. El mismo que pueden ser:</p> <ul style="list-style-type: none"> <li>• Baja: 0-70%</li> <li>• Alta: 71-100%</li> </ul>
<p><b><i>opiniones:</i></b></p>

De tipo varchar, continuo, atributo que permite la recopilación de los criterios más comunes generados del proceso de evaluación docente.

**nota:**

De tipo double, continuo, categoriza el promedio de la nota obtenida por un estudiante en determinada asignatura:

- Mala: 0 – 6.9
- Regular: 7 – 7.5
- Buena: 7.6 – 8.9
- Excelente: 9 – 10

La tabla presenta cada uno de los atributos utilizados en la construcción de estructuras en el proceso de minería, cabe destacar que las dicretizaciones o clasificaciones de porcentajes como: evaluación, calif\_cualitativa entre otros, se los hizo mediante revisiones bibliográficas particularmente de una investigación realizada en la Universidad Nacional Autónoma de México en donde se realizó una investigación similar (Ver Revisión Bibliográfica [20])

TABLA LXXXV: ESTRUCTURA 1

Atributo	Tipo de Dato	Valores
cedula_docente	continuo	
carrera	discreto	<ul style="list-style-type: none"> <li>• Ing.Elect</li> <li>• Ing.Geo y OT</li> <li>• Ing.Sis</li> <li>• Ing.Elect y Telec</li> <li>• Ing.Geo y OT_e</li> <li>• Technol.Elect y CI</li> </ul>
asignatura	continuo	
periodo_academico	discreto	<ul style="list-style-type: none"> <li>• 1</li> <li>• 2</li> <li>• 3</li> </ul>
calif_2_1	continuo	
calif_4_3	continuo	
resultado	continuo	
porcentaje	continuo	
evaluacion	discreto	<ul style="list-style-type: none"> <li>• Criter-Int</li> <li>• Buen-Rend</li> </ul>
cpf_4_3, cpg_4_3, pv_4_3	continuo	
cpf_2_1, cpg_2_1,	continuo	

pv_2_1		
resul_cpf	continuo	
resul_cpg	continuo	
resul_pv	continuo	
porc_cpf, porc_cpg, porc_pv	discreto	<ul style="list-style-type: none"> <li>• regular</li> <li>• buena</li> <li>• excelente</li> </ul>

TABLA LXXXVI: ESTRUCTURA 2,3

<b>Atributo</b>	<b>Tipo de Dato</b>	<b>Valores</b>
cedula_docente	continuo	
carrera	discreto	<ul style="list-style-type: none"> <li>• Ing.Elect</li> <li>• Ing.Geo y OT</li> <li>• Ing.Sis</li> <li>• Ing.Elect y Telec</li> <li>• Ing.Geo y OT_e</li> <li>• Technol.Elect y CI</li> </ul>
calf_4_3	continuo	
calf_2_1	continuo	
muy_alto	discreto	<ul style="list-style-type: none"> <li>• 0</li> <li>• 1</li> </ul>
alto	discreto	<ul style="list-style-type: none"> <li>• 0</li> <li>• 1</li> </ul>
bajo	discreto	<ul style="list-style-type: none"> <li>• 0</li> <li>• 1</li> </ul>
muy_bajo	discreto	<ul style="list-style-type: none"> <li>• 0</li> <li>• 1</li> </ul>
periodo_academico	discreto	<ul style="list-style-type: none"> <li>• 1</li> <li>• 2</li> <li>• 3</li> </ul>
total_calf	continuo	
catedra	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> </ul>
catedra_p_1	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> <li>• A</li> </ul>
catedra_p_2	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> <li>• A</li> </ul>

catedra_p_3	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> <li>• A</li> </ul>
evol_catedra	discreto	<ul style="list-style-type: none"> <li>• Si</li> <li>• No</li> </ul>
cpf_4_3	continuo	
cpf_2_1	continuo	
cpg_4_3	continuo	
cpg_2_1	continuo	
pv_4_3	continuo	
pv_2_1	continuo	
resul_cpf	continuo	
resul_cpg	continuo	
resul_pv	continuo	
catedra_cpf	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> </ul>
catedra_cpg	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> </ul>
catedra_pv	discreto	<ul style="list-style-type: none"> <li>• N</li> <li>• P</li> </ul>

TABLA LXXXVII: ESTRUCTURA 4

Atributo	Tipo de Dato	Valores
calf_2_1	continuo	
calf_4_3	continuo	
calif_cualitativa:	discreto	<ul style="list-style-type: none"> <li>• Baja</li> <li>• Alta</li> </ul>
carrera	discreto	<ul style="list-style-type: none"> <li>• Ing.Elect</li> <li>• Ing.Geo y OT</li> <li>• Ing.Sis</li> <li>• Ing.Elect y Telec</li> <li>• Ing.Geo y OT_e</li> <li>• Tecnol.Elect y CI</li> </ul>
cedula_docente	continuo	
cedula_estudiante	continuo	
resultado	continuo	
periodo_academico	discreto	<ul style="list-style-type: none"> <li>• 1</li> <li>• 2</li> <li>• 3</li> </ul>



porcentaje	continuo	
semestre	continuo	• 1-11
nota	continuo	• 0-10

Como se mencionó anteriormente en esta tarea se continuó con la limpieza de los datos, ya que se requirió hacerlo. Una vez generada la estructura para la minería de datos se optó por eliminar las calificaciones finales de los docentes en todas las carreras, ya que las mismas daban un sesgo de información muy grande, debido a que en todas las carreras los mismos docentes figuran como directivos y par académicos a la vez. Recalcando que el análisis se lo realiza en base a la heteroevaluación del reglamento establecido: es decir la que los estudiantes realizan hacia sus docentes. Con esto se trabajó directamente con las evaluaciones generadas por los estudiantes en los 3 periodos académicos, siendo esto a la vez un aspecto válido ya que las calificaciones de los mismos generan el mayor peso de los resultados finales del proceso de EDD.

Existen investigaciones en donde solo se toma la información pertinente y adecuada que no de un sesgo de información y que pueda incurrir en la mala generación de modelos de minería (Ver Revisión Bibliográfica [20]).

#### **3.1.3.4 Integración de datos**

La integración de los datos se llevó a cabo uniendo la base de datos correspondiente a la EDD, con las notas de los estudiantes en los periodos de estudio en el área seleccionada.

La base de datos de EDD proporcionada por la UTI contiene los registros de 2 Encuestas de Satisfacción Estudiantil y una Encuesta de Desempeño realizadas en 3 periodos académicos: 2011, 2012, 2013, entre evaluaciones, asignaturas, periodos, cuestionarios, contestaciones, docentes y estudiantes, dando un total de 557.411 registros.

Dentro de la integración fue preciso agregar más tablas que ayuden al proceso de análisis por ejemplo:

- Tabla (docente\_evaluacion), con atributos: id\_doc\_period\_acad, cedula, nombres, apellidos

- Tabla (estudiante\_evaluacion), con atributos: id\_est\_period\_acad, cedula, nombres, apellidos.
- Tabla (est\_c) con atributos: *cedula\_docent, cedula\_estudiante, carrera, asignatura, periodo\_academico\_id, respuesta, evaluacion\_id*

Tablas generadas de la misma base de datos, las mismas que fueron útiles en la generación de las estructuras anteriormente descritas.

Una vez terminado la integración de los datos se muestra a continuación (Ver Figura 25) el diagrama de la base de datos final.

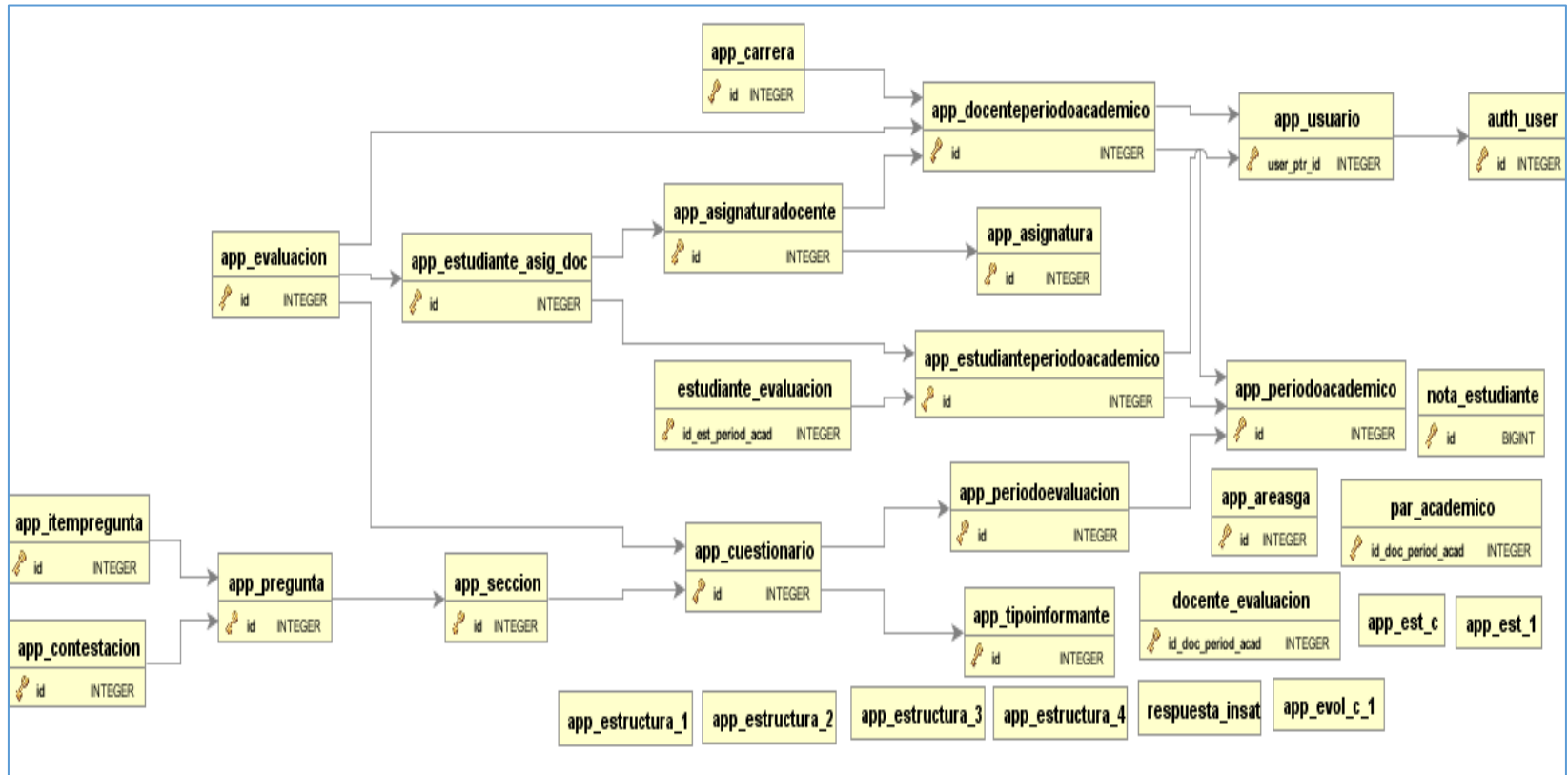


Figura 25: Diagrama de la base de datos final.

### **3.1.4 MODELADO**

Una de las principales ventajas que presenta la metodología CRISP\_DM, es su flexibilidad en cuanto a la transición de sus fases, típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de datos. Por lo tanto, volver a la fase de preparación de datos es a menudo necesario.

En este aspecto para obtener resultados óptimos fue necesario, transformar, seleccionar o estructurar un set de datos para de esta forma generar los modelos de data mining válidos e interpretables descritos en la fase 3.

En esta fase de modelado se seleccionó las técnicas de minería apropiadas para el conjunto de datos provistos, como se mencionó anteriormente se puede aplicar varias técnicas para el mismo problema de minería; por lo tanto se presentan los procesos llevados a cabo y resultados arrojados por cada modelo.

Además la fase de modelado involucra diversos pasos o tareas que permiten llevar a cabo un desarrollo ordenado y objetivo en cada uno de los procesos requeridos.

#### **3.1.4.1 Selección de la técnica de modelado**

Para seleccionar de la técnica de modelado, se debe tener siempre presente los objetivos del negocio, los procesos que se requieren y la disposición y estructura de los datos.

A continuación se describen las técnicas seleccionadas en la generación del o los modelos de minería de datos.

- **Algoritmos de Regresión**

Los algoritmos de regresión permiten determinar la contribución o influencia de varios factores a un resultado. La regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica. Es útil para modelar la probabilidad de un evento en función de otros factores.

Los modelos de regresión logística son modelos en los que se desea conocer la relación entre:

Una variable dependiente cualitativa, (regresión logística binaria o binomial) o con más de dos valores (regresión logística multinomial).

Es decir se busca encontrar la probabilidad  $y \in [0,1]$  de ocurrencia de un evento en base a utilizar regresores  $\mathbf{x} \in \mathbb{R}^n$ , los que son extraídos a partir de los datos. El modelo calcula una función logística que estima la probabilidad de ocurrencia del fenómeno utilizando datos de entrada  $x_i$  con  $i \in \{1, \dots, N\}$  la muestra de  $N$  objetos que se poseen de muestra, y las etiquetas (salidas) reales del modelo  $y_i \in \{0,1\}$  [94].

Se optó por el algoritmo de regresión logística para el modelado, el mismo que se encuentra detallado en el Capítulo II de la Revisión Literaria Recopilación De Técnicas De Minería De Datos.

- **Algoritmos de Clasificación**

Un árbol de clasificación o decisión es una forma de representar el conocimiento obtenido en el proceso de aprendizaje inductivo. Este tipo de algoritmos sirven para el modelado de predicción o descripción de atributos discretos y continuos [20].

Es uno de los métodos de aprendizaje inductivo supervisado no paramétrico más utilizado. Como forma de representación del conocimiento, los árboles de clasificación se destacan por su sencillez, capaces de tomar decisiones con gran precisión [22-26].

Los algoritmos de clasificación elegidos para la construcción de modelos son: C4.5, CHAID, REPTree e ID3, Naive Bayes y descritos en el Capítulo II de la Revisión Literaria: Recopilación De Técnicas De Minería De Datos.

- **Algoritmos Basados en Reglas**

A través de estos algoritmos se puede generar reglas, muchas de las veces son fáciles de interpretar, pero además presentan condiciones que se dan dentro de la base de datos. Los algoritmos probados dentro de la investigación fueron JRip, Ridor, Part, Prism, NNge, no de todos los algoritmos se obtuvo buenos resultados por lo tanto se seleccionó el que mejor resultados presentó, siendo el algoritmo NNge, dando un aplicación y uso sencillo y entendible.

#### **3.1.4.1.1 Salidas del Modelo**

Las técnicas de modelado por lo general requieren que los datos tengan determinadas características. A continuación se describe las salidas del modelado:

##### ***a) Técnica de Modelado***

**Árboles de Clasificación:** se usó para describir el comportamiento de los estudiantes ante la evaluación de desempeño docente en base a las notas obtenidas en una asignatura, cabe recalcar que este algoritmo es un refuerzo del algoritmo de predicción logística aplicado anteriormente, pero también fue útil en la generación de más modelos.

**Regresión Logística:** se usó para predecir la calificación que tendrá un docente en base a la nota que ha tenido un estudiante en determinada asignatura, el grado de influencia de las notas en cada una de las carreras.

**Generación de Reglas: JRip, Naive Bayes, Ridor, Part, Prism, NNge:** estos algoritmos fueron útiles en la generación de reglas, describir o clasificar a los docentes que han tenido una evolución de cátedra, buen rendimiento o criterios de interés en los periodos académicos estudiados unos tomados y otros no, debido a la poca generación de reglas o irrelevancia en las mismas.

##### ***b) Supuestos del modelado***

Se contó con la cantidad de registros y con la calidad requerida por los algoritmos. También con las técnicas y herramientas seleccionadas para cada proceso de modelado.

#### **3.1.4.2 Generación del diseño de pruebas**

La generación del diseño de pruebas consiste en probar la calidad y validez de los resultados arrojados por el modelo que luego se implementará, previo a esto se realizó una serie de ensayos con el fin de probar cada uno de los modelos y determinar el más óptimo para el conjunto de datos de las estructuras generadas, y así observar y analizar el comportamiento de los datos dentro del proceso de evaluación de desempeño docente.

### **3.1.4.3 Construcción de Modelo**

Para la construcción del modelo se debe tener presente que es lo que se requiere analizar, hacia donde se pretende llegar, pero sobre todo se debe poner mucha atención a los patrones de comportamiento que puede generar un modelo en base a los algoritmos aplicados, patrones que a su vez pueden ser conocimiento o factores de interés.

En base a esto y para generar un análisis óptimo del proceso de evaluación de desempeño se aprovechó al máximo los recursos y herramientas disponibles, dando así un enfoque basado en los objetivos y alcance determinado en el presente trabajo de investigación.

Por lo tanto el enfoque de análisis está basado en:

- ✓ Los criterios de interés (*opiniones*) que puede tener el desempeño de un docente en base a la calificación de los estudiantes.
- ✓ La evolución de cátedra que ha tenido un docente en los diferentes periodos académicos.
- ✓ El grado de influencia que tienen las notas de los estudiantes en la calificación de desempeño docente.

#### **3.1.4.3.1 Criterios de Interés existentes en el proceso de evaluación docente.**

Una de las actividades dentro del proceso de evaluación de desempeño es obtener los criterios de interés u opiniones de los estudiantes con respecto al desenvolvimiento de un docente en su espacio de cátedra, esto con el fin de aportar al plan de mejora individual de cada docente, de la misma forma se obtiene la evaluación de un docente que ha tenido un buen rendimiento en un periodo de evaluación.

En todos los casos de modelado se evalúa primero el rendimiento del modelo, y luego explicarlo o predecir la variable de interés (*label*). De igual manera para la validación cruzada se utiliza 7 subconjuntos para evaluar la generación y aplicación del modelo en los datos (Ver anexo 7).

Para la realización de las siguientes pruebas se hizo uso de una estructura que contiene el DNI de un docente y las asignaturas dictadas dentro de un periodo académico en una carrera.

### 3.1.4.3.1.1 Clasificación mediante ID3

Mediante el árbol de clasificación se pudo observar de manera gráfica y estructurada la existencia de criterios de interés o el rendimiento de un docente en el proceso de evaluación. Los parámetros que se establecieron fueron el número mínimo de divisiones que se pueden dar por cada nodo (*minimal size for split=3*), el tamaño mínimo de cada hoja (*minimal leaf size=2*), la ganancia mínima (*minimal gain=0.6*), y (*accuracy*) que maximiza la precisión de todo el árbol, *en la validación cruzada se seleccionó 7 subconjuntos para evaluar el modelo generado.*

Además se establecieron roles para la generación del modelo; *cedula\_docent* como id, y *evaluacion* como label. Cabe destacar que las condiciones generadas por el árbol de clasificación son basadas en el atributo label es decir si existen o no criterios de interés de los estudiantes hacia un docente, en un periodo académico, en una asignatura y dentro de una carrera, o si existen opiniones que recalquen la excelencia de cátedra que posee un docente.

Los resultados dados por el algoritmo ID3 son los siguientes (Ver Tabla LXXXVIII):

TABLA LXXXVIII: RENDIMIENTO, MODELO, ALGORITMO ID3

<b>Algoritmo ID3</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	100%	97.25%
Valores incorrectos en la clasificación	0.0%	2.75%

En la tabla se muestra que para el desempeño del modelo existe un 100% de clasificación y para la validación cruzada existe un 97.25% de instancias correctamente clasificadas.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 26):



accuracy: 97.25% +/- 1.07% (mikro: 97.25%)			
	true Criter-Int	true Buen-Rend	class precision
pred. Criter-Int	36	9	80.00%
pred. Buen-Rend	6	494	98.80%
class recall	85.71%	98.21%	

Figura 26: Matriz de confusión árbol ID3.

Para la descripción de la evaluación en la existencia de criterios de interés o buen rendimiento en la matriz de confusión se puede observar, que para la existencia de Criterios de Interés existe un 80% de precisión y para Buen Rendimiento un 98.80%. En la siguiente figura (Ver Figura 27) se presenta parte del árbol generado:

```

Tree
carrer = Ing.Elect y Telec
|   porc_cpg = buena
|   |   porc_cpf = buena: Buen-Rend {Criter-Int=0, Buen-Rend=8}
|   |   |   porc_cpf = regular
|   |   |   |   periodo_acad_id = 1: Buen-Rend {Criter-Int=0, Buen-Rend=1}
|   |   |   |   |   periodo_acad_id = 3: Criter-Int {Criter-Int=1, Buen-Rend=0}
|   |   |   |   |   |   porc_cpg = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=85}
|   |   |   |   |   |   |   porc_cpg = regular
|   |   |   |   |   |   |   |   porc_cpf = buena: Criter-Int {Criter-Int=1, Buen-Rend=0}
|   |   |   |   |   |   |   |   |   porc_cpf = regular
|   |   |   |   |   |   |   |   |   |   porc_pv = buena: Criter-Int {Criter-Int=4, Buen-Rend=0}
|   |   |   |   |   |   |   |   |   |   |   porc_pv = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=1}
|   |   |   |   |   |   |   |   |   |   |   |   porc_pv = regular: Criter-Int {Criter-Int=6, Buen-Rend=0}
carrer = Ing.Electm
|   porc_cpg = buena
|   |   porc_pv = buena: Buen-Rend {Criter-Int=0, Buen-Rend=6}
|   |   |   porc_pv = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=9}
|   |   |   |   porc_pv = regular: Criter-Int {Criter-Int=1, Buen-Rend=0}
|   |   |   |   |   porc_cpg = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=102}
|   |   |   |   |   |   porc_cpg = regular: Criter-Int {Criter-Int=12, Buen-Rend=0}

```

Figura 27: Árbol de clasificación ID3

El árbol toma nodo raíz las carreras del área de la energía y hace un desglose hacia los porcentajes de calificación por componente de evaluación siendo la capacidad pedagógica la de mayor. El árbol clasifica las variables más relevantes tomando en cuenta la clase o label a predecir.

Es decir si un docente pertenece a determinada carrera y posee un porcentaje de calificación regular en cuanto a su capacidad pedagógica entonces probablemente existirán criterios de interés, pero si este porcentaje es alto o excelente entonces tendrá un buen rendimiento que deberían tomarse en cuenta al generar un plan de mejoras hacia un docente.

### 3.1.4.3.1.2 Clasificación mediante Árbol CHAID

Para el modelo propuesto se fijó los siguientes parámetros:

- minimal size for split= 4
- minimal leaf size= 2
- minimal gain=0.6
- *maximal deph*=10
- *confidence*=0.5
- *number of prepruning*=3

Esto es la longitud mínima para el parámetro de división en cada nodo, el subconjunto o tamaño mínimo de un nodo hoja, la ganancia mínima de un nodo antes de su división, la profundidad máxima o tamaño del árbol de decisión, el nivel de confianza utilizado para el cálculo de error, número de nodos probados antes de una división (por defecto), en la validación cruzada se seleccionó 7 subconjuntos para evaluar el modelo generado.

Al activar la opción de *no prepruning*, se obtiene un árbol sin una pre poda, *no pruning* (por defecto) al finalizar el modelo un árbol sin podar para poder obtener una clasificación óptima.

Los resultados en el modelo son los siguientes (Ver Tabla LXXXIX):

TABLA LXXXIX: RENDIMIENTO, MODELO CHAID

<b>Algoritmo CHAID</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	99.82%	98.90%
Valores incorrectos en la clasificación	0.18%	1.10%

En la tabla se muestra que para el desempeño del modelo o algoritmo se da un 98.82% de las instancia correctamente clasificadas, mientras que en la validación cruzada esta baja dando un 98.90%.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 28):

accuracy: 98.90% +/- 1.27% (mikro: 98.90%)			
	true Criter-Int	true Buen-Rend	class precision
pred. Criter-Int	40	4	90.91%
pred. Buen-Rend	2	499	99.60%
class recall	95.24%	99.20%	

Figura 28: Matriz de confusión CHAID

Para la predicción en la existencia de criterios de interés se establece una precisión de 90.91% mientras que para un Buen rendimiento existe una precisión de 99.60%. En la (Figura 29) se presenta un segmento del árbol generado:

```

Tree
periodo_acad_id = 1
| porc_cpf = buena: Buen-Rend {Criter-Int=0, Buen-Rend=30}
| porc_cpf = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=142}
| porc_cpf = regular
| | porc_cpg = buena
| | | porc_pv = buena: Buen-Rend {Criter-Int=0, Buen-Rend=1}
| | | porc_pv = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=1}
| | | porc_pv = regular: Criter-Int {Criter-Int=1, Buen-Rend=0}
| | | porc_cpg = regular
| | | porc_pv = buena
| | | | carrer = Ing.Elect y Telec: Criter-Int {Criter-Int=1, Buen-Rend=0}
| | | | carrer = Ing.Electm: Criter-Int {Criter-Int=2, Buen-Rend=0}
| | | | carrer = Ing.Geo y OT: Criter-Int {Criter-Int=1, Buen-Rend=0}
| | | | carrer = Ing.Sis: Buen-Rend {Criter-Int=0, Buen-Rend=1}
| | | | porc_pv = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=1}
| | | | porc_pv = regular: Criter-Int {Criter-Int=9, Buen-Rend=0}
periodo_acad_id = 2
| porc_cpf = buena: Buen-Rend {Criter-Int=1, Buen-Rend=12}
| porc_cpf = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=143}
| porc_cpf = regular
| | porc_cpg = buena: Buen-Rend {Criter-Int=0, Buen-Rend=5}
| | porc_cpg = regular: Criter-Int {Criter-Int=15, Buen-Rend=0}

```

Figura 29: Árbol de clasificación CHAID.

El árbol toma como nodo raíz los periodos académicos y luego hace una comparación del porcentaje de calificación por componentes en la evaluación, donde se puede evidenciar donde y cuanto existe un buen rendimiento docente y que carreras presentan criterios de interés.

### 3.1.4.3.1.3 Clasificación mediante Árbol C4.5.

Para el modelo propuesto se fijó los siguientes parámetros:

- minimal size for split= 4
- minimal leaf size= 2
- minimal gain=0.6
- *maximal dep*h=10

- *confidence=0.5*
- *number of prepruning=3*
- *accuracy*

Esto es la longitud mínima para el parámetro de división en cada nodo, el subconjunto o tamaño mínimo de un nodo hoja, la ganancia mínima de un nodo antes de su división, la profundidad máxima o tamaño del árbol de decisión, el nivel de confianza utilizado para el cálculo de error, número de nodos probados antes de una división (por defecto) y la precisión de todo el árbol, en la validación cruzada se seleccionó 7 subconjuntos para evaluar el modelo generado.

Al activar las opciones de *no prepruning* y *prepruning*, se obtiene un árbol sin una pre poda y al finalizar el modelo un árbol sin podar para poder obtener una clasificación óptima. Los resultados en el modelo son los siguientes (Ver Tabla XC):

TABLA XC: RENDIMIENTO, MODELO C4.5

<b>Algoritmo C4.5</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	96.70%	94.32%
Valores incorrectos en la clasificación	3.30%	5.68%

En la tabla se muestra que para el rendimiento del modelo se da un 96.70% de las instancia correctamente clasificadas mientras que en la validación cruzada esta baja dando un 94.32%.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 30):

<b>accuracy: 94.32% +/- 2.72% (mikro: 94.31%)</b>			
	true Criter-Int	true Buen-Rend	class precision
pred. Criter-Int	19	8	70.37%
pred. Buen-Rend	23	495	95.56%
class recall	45.24%	98.41%	

Figura 30: Matriz de confusión C4.5

Para la predicción en la existencia de criterios de interés se establece una precisión de 70.37% mientras que para un Buen rendimiento existe una precisión de 95.56%. En la (Figura 31) se presenta un segmento del árbol generado:

```

Tree

carrer = Ing.Elect y Telec
|   porc_cpg = buena
|   |   porc_cpf = buena: Buen-Rend {Criter-Int=0, Buen-Rend=8}
|   |   porc_cpf = regular
|   |   |   periodo_acad_id = 1: Buen-Rend {Criter-Int=0, Buen-Rend=1}
|   |   |   periodo_acad_id = 3: Criter-Int {Criter-Int=1, Buen-Rend=0}
|   |   porc_cpg = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=85}
|   |   porc_cpg = regular: Criter-Int {Criter-Int=11, Buen-Rend=1}
carrer = Ing.Electm: Buen-Rend {Criter-Int=13, Buen-Rend=117}
carrer = Ing.Geo y OT: Buen-Rend {Criter-Int=4, Buen-Rend=130}
carrer = Ing.Geo y OT_e
|   periodo_acad_id = 1
|   |   porc_cpf = excelente: Buen-Rend {Criter-Int=0, Buen-Rend=5}
|   |   porc_cpf = regular: Criter-Int {Criter-Int=2, Buen-Rend=0}
|   periodo_acad_id = 2: Buen-Rend {Criter-Int=0, Buen-Rend=9}
carrer = Ing.Sis
|   porc_cpg = buena
|   |   porc_pv = buena
|   |   |   periodo_acad_id = 1: Buen-Rend {Criter-Int=0, Buen-Rend=3}
|   |   |   periodo_acad_id = 2: Buen-Rend {Criter-Int=0, Buen-Rend=4}
|   |   |   periodo_acad_id = 3

```

Figura 31: Árbol de clasificación C4.5

El árbol toma como nodo raíz las carreras y luego hace una comparación del porcentaje de calificación por componentes y periodo académico, donde se puede evidenciar donde y cuanto existe un buen rendimiento docente y que carreras presentan criterios de interés.

### 3.1.4.3.1.4 Generación de Reglas mediante el algoritmo W-PART

Se fijaron los siguientes parámetros:

- C: 0.6 confianza en la generación de reglas
- M: número mínimo de objetos por hoja (por defecto 2)
- R: reduce la poda error
- N: se establece el número de pliegues para reducir el error de poda
- B: denota divisiones binarias (falso por defecto).
- U: lista de decisión sin podar (falso por defecto).
- Q: Semilla de datos aleatorios (por defecto 1).

Los resultados en el modelo son los siguientes (Ver Tabla XCI):

TABLA XCI: RENDIMIENTO, MODELO W-PART

Algoritmo W-PART		
	Descripción	
	Desempeño Modelo	Validación Cruzada
Valores correctamente clasificados	99.82%	99.27%
Valores incorrectos en la clasificación	0.18%	0.73%

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 32):

accuracy: 99.27% +/- 1.35% (mikro: 99.27%)			
	true Criter-Int	true Buen-Rend	class precision
pred. Criter-Int	39	1	97.50%
pred. Buen-Rend	3	502	99.41%
class recall	92.86%	99.80%	

Figura 32: Matriz de confusión W-PART

Para la predicción en la existencia de criterios de interés se establece una precisión de 97.50% mientras que para un Buen rendimiento existe una precisión de 99.41%. En la (Ver Figura 33) se presentan las reglas generadas por el algoritmo:

```

W-PART

PART decision list
-----

porc_pv = excelente: Buen-Rend (471.0)

porc_cpg = regular AND
carrer = Ing.Electm: Criter-Int (12.0)

porc_cpg = regular AND
periodo_acad_id = 2: Criter-Int (15.0)

porc_pv = buena AND
porc_cpf = buena: Buen-Rend (23.0)

porc_cpf = regular AND
periodo_acad_id = 3: Criter-Int (9.0)
    
```

Figura 33: Reglas mediante W-PART

La regla más relevante denota que existen 471 docentes con una evaluación en un Buen Rendimiento cuando estos han tenido una práctica de valores excelente. El total de reglas

generadas son 9 tomando como referencia ya sea los componentes, carreras y periodos de evaluación.

### 3.1.4.3.1.5 Generación de Reglas mediante el algoritmo W-NNge

En el presente algoritmo se tomó al igual que los anteriores 7 subconjuntos para evaluar el modelo generado, además de los siguientes parámetros:

G: Número de intentos de generalización =0.5. Por defecto

I: Número para el cálculo de la información mutua 0.5. Por defecto

Los resultados en el modelo son los siguientes (Ver Tabla XCII):

TABLA XCII: RENDIMIENTO, MODELO W-NNge

Algoritmo W-NNge		
	Descripción	
	Desempeño Modelo	Validación Cruzada
Valores correctamente clasificados	100%	98.89%
Valores incorrectos en la clasificación	0.00%	1.10%

La matriz de confusión arrojó los siguientes resultados (Ver Figura 34):

```

=== Confusion Matrix ===
      a  b  <-- classified as
40   2 |  a = Criter-Int
 4 499 |  b = Buen-Rend
    
```

Figura 34: Matriz de confusión W-NNge

Para la existencia de criterios de interés se establece un 7.33% mientras que para un Buen rendimiento existe una precisión de 91.55%. En la (Ver Figura 35) se presentan las reglas generadas por el algoritmo:

```

NNGE classifier

Rules generated :
class Buen-Rend IF : periodo_acad_id in {2,3,1} ^ porc_cpf in {buena,excelente} ^ porc_cpg in {buena,excelente} ^ porc_pv in {buena,excelente}
class Criter-Int IF : periodo_acad_id in {2,3,1} ^ porc_cpf in {regular} ^ porc_cpg in {regular,buena} ^ porc_pv in {regular} ^ carrer in {Ing.
class Buen-Rend IF : periodo_acad_id in {3} ^ porc_cpf in {regular} ^ porc_cpg in {buena} ^ porc_pv in {excelente} ^ carrer in {Ing.Sis} (1)
class Buen-Rend IF : periodo_acad_id in {1} ^ porc_cpf in {regular} ^ porc_cpg in {regular} ^ porc_pv in {buena} ^ carrer in {Ing.Sis} (1)
class Buen-Rend IF : periodo_acad_id in {1} ^ porc_cpf in {regular} ^ porc_cpg in {regular} ^ porc_pv in {excelente} ^ carrer in {Ing.Elect y T
class Buen-Rend IF : periodo_acad_id in {2,1} ^ porc_cpf in {regular} ^ porc_cpg in {buena} ^ porc_pv in {buena,excelente} ^ carrer in {Ing.Ele
class Criter-Int IF : periodo_acad_id in {2,3} ^ porc_cpf in {regular} ^ porc_cpg in {regular} ^ porc_pv in {buena} ^ carrer in {Ing.Sis} (3)
class Criter-Int IF : periodo_acad_id in {3} ^ porc_cpf in {regular} ^ porc_cpg in {buena} ^ porc_pv in {buena} ^ carrer in {Ing.Elect y Telec,
class Criter-Int IF : periodo_acad_id in {2,3,1} ^ porc_cpf in {regular,buena} ^ porc_cpg in {regular} ^ porc_pv in {buena} ^ carrer in {Ing.El

```

Figura 35: Reglas mediante W-NNge

### Reglas relevantes:

class Buen-Rend IF : periodo\_acad\_id in {2,3,1} ^ porc\_cpf in {buena,excelente} ^ porc\_cpg in {buena,excelente} ^ porc\_pv in {buena,excelente} ^ carrer in {Ing.Electm,Ing.Geo y OT,Ing.Elect y Telec,Ing.Sis,Ing.Geo y OT\_e,Tecnol.Elect y CI} (493)

class Criter-Int IF: periodo\_acad\_id in {2, 3,1} ^ porc\_cpf in {regular, buena} ^ porc\_cpg in {regular} ^ porc\_pv in {buena} ^ carrer in {Ing.Electm, Ing.Geo y OT, Ing.Elect y Telec} (9)

La regla más relevante denota que existen 493 docentes con una evaluación en un Buen Rendimiento, las reglas generadas son 9 de las cuales 5 son de Buen Rendimiento y 4 para Criterios Interés cuando, si bien es cierto las reglas son lógicas y nos dan una pauta de que los modelos generados están bien enmarcados el presente algoritmo no se lo toma debido, a que las reglas no son suficientes.

#### 3.1.4.3.1.6 Modelos W-Ridor, W-ConjunctiveRule

A más de los modelos presentados se probó con más algoritmos como: W-Ridor, W-ConjunctiveRule, W-OneR, entre otros, los mismos que no dieron buenos resultados. A continuación se presenta 2 de los modelos adicionales:

- **W-Ridor**

Con el presente algoritmo se establece una sola condición en la existencia de criterios de interés, y 5 reglas dentro de un buen rendimiento por lo tanto no se estima conveniente la aplicación ni evaluación del presente algoritmo dentro del análisis (Ver Figura 36).



```

W-Ridor

Ripple Down Rule Learner(Ridor) rules
-----

evaluacion = Criter-Int (545.0/503.0)
  Except (porc_cpg = excelente) => evaluacion = Buen-Rend (296.0/0.0) [149.0/0.0]
  Except (porc_cpg = buena) and (porc_cpf = buena) => evaluacion = Buen-Rend (27.0/0.0) [14.0/0.0]
  Except (porc_cpg = buena) and (porc_cpf = excelente) => evaluacion = Buen-Rend (6.0/0.0) [1.0/0.0]
  Except (porc_cpg = buena) and (periodo_acad_id = 2) => evaluacion = Buen-Rend (3.0/0.0) [2.0/0.0]
  Except (porc_cpg = buena) => evaluacion = Buen-Rend (5.0/3.0) [1.0/0.0]

Total number of rules (incl. the default rule): 6

```

Figura 36: Reglas mediante W-Ridor

- **W-ConjunctiveRule**

Este algoritmo arrojó solo una regla que especifica la existencia de criterios de interés cuando dentro de la evaluación existe una capacidad profesional regular, por lo tanto no se lo toma en cuenta dentro del análisis (Ver Figura 37).

```

W-ConjunctiveRule

Single conjunctive rule learner:
-----

(porc_cpf = regular) => evaluacion = Criter-Int

Class distributions:
Covered by the rule:
Criter-Int      Buen-Rend
0.848485        0.151515

Not covered by the rule:
Criter-Int      Buen-Rend
0                1

```

Figura 37: Reglas mediante W-ConjunctiveRule

Cabe recalcar que existen diversos algoritmos a la aplicación de un problema de minería de datos y llevaría demasiado tiempo estudiarlos todos. Por lo tanto tomando como referencia la revisión bibliográfica, aplicación y rendimiento de los mismos se opta por los algoritmos con mejores resultados.

### 3.1.4.3.1.7 Obtención de Opiniones:

Una de las formas de obtención de opiniones es mediante la técnica de Texmining, una de las herramientas que proporciona el análisis de textos aplicando tex-mining es Many Eyes en el uso de estadísticas y número de ocurrencias, la cual permitió una visualización de las opiniones más relevantes desde diferentes perspectivas de análisis cabe destacar



Los resultados dados por el algoritmo Naive Bayes fueron los siguientes (Ver Tabla XCIII):

**TABLA XCIII: RENDIMIENTO, MODELO, ALGORITMO NAIVE BAYES**

<b>Algoritmo Naive Bayes</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	93.05%	93.05%
Valores incorrectos en la clasificación	6.95%	6.95%

En la tabla anterior se muestra que tanto para el desempeño del modelo como para la validación cruzada existe más del 90% de instancias correctamente clasificadas.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 39):

accuracy: 93.05%			
	true Si	true No	class precision
pred. Si	163	6	96.45%
pred. No	12	78	86.67%
class recall	93.14%	92.86%	

Figura 39: Matriz de confusión Naive Bayes

En la matriz de confusión se puede observar, que para la clase Si existe un acierto mayor al 90% y para la clase No un 86.67%. En la (Figura 40) se presenta la clasificación generada:

```

SimpleDistribution

Distribution model for label attribute evol_catedra

Class Si (0.676)
8 distributions

Class No (0.324)
8 distributions
    
```

Figura 40: Clasificación Naive Bayes

La clasificación se dio con una distribución de 0.324 para la clase No y 0.676 para la clase Sí. Existe una buena clasificación por medio del presente algoritmo, además de presentar

un buen rendimiento del modelo y validación cruzada, presenta una buena distribución de las clases predictivas o label.

### 3.1.4.3.2 Clasificación mediante ID3

Mediante el árbol de clasificación se pudo observar de manera gráfica y estructurada la existencia o no de una evolución de cátedra en los periodos académicos de estudio. Los parámetros que se establecieron fueron el número mínimo de divisiones que se pueden dar por cada nodo (*minimal size for split=3*), el tamaño mínimo de cada hoja (*minimal leaf size=3*), la ganancia mínima (*minimal gain=0.7*), y (*accuracy*) que maximiza la precisión de todo el árbol, en la validación cruzada se seleccionó 7 subconjuntos para evaluar el modelo generado.

Además se establecieron roles para la generación del modelo; *cedula\_docent* como id, y *evol\_catedra* como label. Cabe destacar que las condiciones generadas por el árbol de clasificación son basadas en el atributo label es decir si existen o no evolución de cátedra de un docente, en varios periodos académicos.

Los resultados dados por el algoritmo ID3 son los siguientes (Ver Tabla XCIV)

TABLA XCIV: RENDIMIENTO, MODELO, ALGORITMO ID3

<b>Algoritmo ID3</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	97.30%	92.28%
Valores incorrectos en la clasificación	2.70%	7.72%

En la tabla se muestra que tanto para el desempeño del modelo como para la validación cruzada existe más del 90% de precisión.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 41):

accuracy: 92.28% +/- 5.86% (mikro: 92.28%)			
	true Si	true No	class precision
pred. Si	166	11	93.79%
pred. No	9	73	89.02%
class recall	94.86%	86.90%	

Figura 41: Matriz de confusión árbol ID3.

La matriz de confusión describe la evolución de cátedra como si y no, dando una precisión de clasificación de las clases de 93.79% y 89.02% respectivamente. En la (Figura 42) se presenta el árbol generado:

```

Tree

carrera = Ing.Elect y Telec
|  catedra_cpg = N
|  |  catedra_cpf = N
|  |  |  catedra_pv = N: No {Si=0, No=16}
|  |  |  catedra_pv = P: Si {Si=4, No=0}
|  |  catedra_cpf = P: No {Si=0, No=4}
|  |  catedra_cpg = P: Si {Si=18, No=0}
carrera = Ing.Electm
|  muy_alta = false: No {Si=0, No=9}
|  muy_alta = true
|  |  baja = false
|  |  |  catedra_cpg = N
|  |  |  |  catedra_cpf = N
|  |  |  |  |  catedra_pv = N
|  |  |  |  |  |  alta = false
|  |  |  |  |  |  |  muy_baja = false: No {Si=2, No=7}
|  |  |  |  |  |  |  muy_baja = true: Si {Si=1, No=0}
|  |  |  |  |  |  |  alta = true: No {Si=0, No=2}
|  |  |  |  |  |  |  catedra_pv = P: No {Si=0, No=1}
|  |  |  |  |  |  |  catedra_cpf = P: Si {Si=3, No=0}
|  |  |  |  |  |  |  catedra_cpg = P: Si {Si=47, No=0}
|  |  |  |  |  |  |  baja = true: No {Si=0, No=2}
carrera = Ing.Geo y OT
|  baja = false

```

Figura 42: Árbol de clasificación ID3

El árbol desglosa la información tomando como raíz cada una de las carreras del área de la energía, clasificando los datos según los componentes de evaluación y haciendo comparaciones con las calificaciones totales de cada docente, realizando las validaciones correspondientes según sean o no verdaderas las condiciones especificadas por el algoritmo.

### 3.1.4.3.2.3 Clasificación mediante W-RETree

En el siguiente algoritmo se establecieron los siguientes parámetros M: establece el número mínimo de casos por hoja (por defecto 2), V: ajuste de proporción

(predeterminado  $1e-3$ ), N: número de pliegues para la reducción de error (por defecto 3), S: semilla de datos aleatorios (por defecto 1), P: no poda, por defecto: false, L: profundidad máxima del árbol (por defecto -1).

Los resultados dados por el algoritmo de W-RETree se presentan en la siguiente tabla (Ver Tabla XCV):

TABLA XCV: RENDIMIENTO, MODELO, ALGORITMO BASADO EN W-RETREE

<b>Algoritmo W-RETree</b>		
	<b>Descripción</b>	
	<b>Desempeño Algoritmo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	96.14%	94.21%
Valores incorrectos en la clasificación	3.86%	5.79%

En la tabla se muestra que para la validación cruzada baja el porcentaje de precisión a un 94.21% con un 5.79% de instancias mal clasificadas.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 43):

accuracy: 94.21% +/- 5.29% (mikro: 94.21%)			
	true Si	true No	class precision
pred. Si	168	8	95.45%
pred. No	7	76	91.57%
class recall	96.00%	90.48%	

Figura 43: Matriz de confusión W-RETree

En la matriz de confusión se puede observar, que para la totalidad de los datos existe una clasificación mayor al 90%, para la clase No un 91.57% y un 95.45% para la clase Sí. En la (Figura 44) se presentan el árbol generado:

```

W-REPTree
REPTree
=====
catedra_cpg = P : Si (147/0) [0/0]
catedra_cpg = N
|  catedra_pv = P
|  |  carrera = Ing.Elect y Telec
|  |  |  catedra_cpf = P : No (4/0) [0/0]
|  |  |  catedra_cpf = N : Si (4/0) [0/0]
|  |  |  carrera = Ing.Sis : Si (8/0) [0/0]
|  |  |  carrera = Ing.Electm : Si (4/1) [0/0]
|  |  |  carrera = Ing.Geo y OT : Si (3/1) [0/0]
|  |  |  carrera = Ing.Geo y OT_e : Si (0/0) [0/0]
|  |  |  carrera = Tecnol.Elect y CI : Si (1/0) [0/0]
|  |  catedra_pv = N
|  |  |  carrera = Ing.Elect y Telec : No (16/0) [0/0]
|  |  |  carrera = Ing.Sis : No (33/0) [0/0]
|  |  |  carrera = Ing.Electm
|  |  |  |  muy_alta = false : No (9/0) [0/0]
|  |  |  |  muy_alta = true
|  |  |  |  |  alta = false

```

Figura 44: Algoritmo W-REPTree

La clasificación del árbol toma como nodo inicio la calificación de cátedra por componente de evaluación, para posteriormente hacer un desglose por carreras estableciendo las condiciones en las que un docente tiene o no una evolución de cátedra.

### 3.1.4.3.2.4 Reglas mediante W-NNge

En el presente algoritmo se tomó al igual que los anteriores 7 subconjuntos para evaluar el modelo generado, además de los siguientes parámetros:

G: Número de intentos de generalización =0.5. Por defecto

I: Número para el cálculo de la información mutua 0.5. Por defecto

Los resultados dados por el algoritmo de W-NNge se presentan en la siguiente tabla (Ver Tabla XCVI):

**TABLA XCVI: RENDIMIENTO, MODELO, ALGORITMO BASADO EN W-NNGE**

<b>Algoritmo W-NNge</b>		
	<b>Descripción</b>	
	<b>Desempeño Algoritmo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	96.91%	94.21%
Valores incorrectos en la clasificación	3.08%	5.79%

En la tabla se muestra que para la validación cruzada baja el porcentaje de precisión a un 94.21% con un 5.79% de instancias mal clasificadas.

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 45):

```

=== Confusion Matrix ===
      a  b  <-- classified as
168  7  |  a = Si
  8 76  |  b = No
  
```

Figura 45: Matriz de confusión W-NNge

En la matriz de confusión se puede observar, que para la clase No existe un 29.34% y un 64.86% para la clase Sí. En la (Figura 46) se presentan las reglas generadas:

```

NNGE classifier
Rules generated :
class Si IF : muy_alta in {false,true} ^ alta in {false,true} ^ baja in {false} ^ muy_baja in {false,true} ^ catedra_cpf in {P} ^ catedra_cpg in {P} ^ catedra_pv in {P,N} ^ carrera in
class No IF : muy_alta in {false,true} ^ alta in {false,true} ^ baja in {false,true} ^ muy_baja in {false,true} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Geo y OT}
class No IF : muy_alta in {true} ^ alta in {false} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Geo y OT}
class No IF : muy_alta in {false} ^ alta in {true} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Geo y OT}
class No IF : muy_alta in {true} ^ alta in {false} ^ baja in {true} ^ muy_baja in {true} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Electra} (2)
class No IF : muy_alta in {true} ^ alta in {true} ^ baja in {false} ^ muy_baja in {true} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Electra} (1)
class No IF : muy_alta in {true} ^ alta in {false,true} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {P,N} ^ carrera in {Ing.Electra}
class Si IF : muy_alta in {true} ^ alta in {false} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Geo y OT_e}
class Si IF : muy_alta in {true} ^ alta in {true} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Geo y OT} (1)
class Si IF : muy_alta in {true} ^ alta in {false} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {N} ^ catedra_cpg in {N} ^ catedra_pv in {P} ^ carrera in {Ing.Geo y OT}
class Si IF : muy_alta in {true} ^ alta in {false,true} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {P} ^ catedra_cpg in {N} ^ catedra_pv in {P} ^ carrera in {Ing.Sis,Ing}
class Si IF : muy_alta in {true} ^ alta in {false,true} ^ baja in {false} ^ muy_baja in {false,true} ^ catedra_cpf in {N} ^ catedra_cpg in {P,N} ^ catedra_pv in {P} ^ carrera in {Ing.}
class No IF : muy_alta in {true} ^ alta in {false,true} ^ baja in {false} ^ muy_baja in {false} ^ catedra_cpf in {P} ^ catedra_cpg in {N} ^ catedra_pv in {N} ^ carrera in {Ing.Electra}
  
```

Figura 46: Algoritmo W-NNge

Las reglas generadas por el algoritmo son 10 para la clase Si y 8 para la clase No dando un total de 18 reglas de las cuales se tomó las más relevantes, la más relevante denota que existen 134 docentes en todas las carreras, con una evolución de cátedra, que tienen su base en la capacidad profesional y pedagógica. Los docentes sin evolución son 63 de todas las carreras y negativa en las capacidades evaluadas.

A más de los modelos presentados se probó con más algoritmos como: W-JRip, W-OneR, W-PRISM, entre otros, los mismos que no dieron buenos resultados. A continuación se presenta uno de los modelos adicionales:



```
W-JRip

JRIP rules:
=====

(catedra_pv = N) and (carrera = Ing.Sis) => evol_catedra=No (33.0/0.0)
(catedra_pv = N) and (muy_alta = false) and (alta = true) => evol_catedra=No (14.0/0.0)
(catedra_cpg = N) and (catedra_pv = N) and (carrera = Ing.Elect y Telec) => evol_catedra=No (11.0/0.0)
(catedra_cpg = N) and (catedra_pv = N) and (carrera = Ing.Electm) => evol_catedra=No (17.0/3.0)
(catedra_cpg = N) and (catedra_pv = N) => evol_catedra=No (13.0/7.0)
(catedra_cpg = N) => evol_catedra=No (24.0/18.0)
=> evol_catedra=Si (147.0/0.0)

Number of Rules : 7
```

Figura 47: Generación de reglas W-Jrip

Las reglas generadas por el algoritmo si bien son muy acertadas y permiten considerar las condiciones para que un docente no tenga una evolución de cátedra, estas no toman todas las condiciones posibles dentro de los datos incidiendo en la generación de escasas reglas, dando solo una regla para la evolución de cátedra. El total de reglas generadas son 7.

### 3.1.4.3.3 Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.

En este punto se realizó la construcción de modelos orientados a clasificar, explicar y predecir el comportamiento de un estudiante al momento de evaluar el desempeño de un docente en base a la nota que ha tenido en una unidad.

En todos los casos de modelado se toma un 50% de los datos como muestra para evaluar el rendimiento del modelo, y luego explicarlo o predecir la variable de interés (*label*). De igual manera para la validación cruzada se utiliza 7 subconjuntos para evaluar la generación y aplicación del modelo en los datos (Ver anexo 7).

#### 3.1.4.3.3.1 Regresión logística

En este caso primero se observa cómo se comporta el modelo, para posteriormente aplicarlo a otro conjunto de datos y predecir la calificación de un docente. El proceso de modelamiento involucra la definición de una variable explicativa o *label*, que servirá para predecir en este caso la calificación de un docente en base a la nota de un estudiante.

Se toma el 50% para evaluar el rendimiento del algoritmo o modelo y la totalidad de los registros para la validación cruzada.

Los resultados dados por el algoritmo son los siguientes (Ver Tabla XCVII):

**TABLA XCVII: RENDIMIENTO, MODELO DE REGRESIÓN LOGÍSTICA**

<b>Algoritmo de Regresión Logística</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	89.26%	88.74%
Valores incorrectos en la clasificación	10.74%	11.26%

A continuación se presenta la matriz de confusión para el presente modelo (Ver Figura 48):

<b>accuracy: 88.74% +/- 0.46% (mikro: 88.74%)</b>			
	true Baja	true Alta	class precision
pred. Baja	903	48	94.95%
pred. Alta	1248	9307	88.18%
class recall	41.98%	99.49%	

Figura 48: Matriz de confusión algoritmo de regresión logística

En la matriz generada se puede observar que la precisión de predicción para cada categoría de calificación es del 94.95% para Baja y 88.18% para Alta. Cabe destacar que el modelo generado por cada algoritmo dependerá de la calidad de los datos, es decir calidad de los mismos influye en el rendimiento del proceso [109], en este caso en específico existen valores nulos debido a la combinación de fuentes de datos diferentes, en cuanto a las notas de los estudiantes que provienen de una base de datos distinta a la de evaluación docente.

Por lo tanto esta baja del modelo con relación a la de modelos anteriores puede deberse a la existencia de valores nulos en las notas de algunos estudiantes [109]. A más de la matriz de confusión el algoritmo genera las reglas de regresión logística en dicho modelo (Ver Figura 49):

W-Logistic	
Logistic Regression with ridge parameter of 1.0E-8	
Coefficients...	
	Class
Variable	Baja
=====	
nota	-1.0463
Intercept	6.3911

Figura 49: Modelo regresión logística

El modelo describe que a mayor nota de un estudiante en un asignatura, menor será la probabilidad que un docente tenga una calificación Baja.

El modelo a más de dar una clasificación alta, también presenta comportamientos de interés que se tomarán en cuenta al momento de evaluar los resultados.

En la siguiente figura (ver figura 50) se puede observar una breve descripción de los datos en el modelo:

### Análisis Gráfico

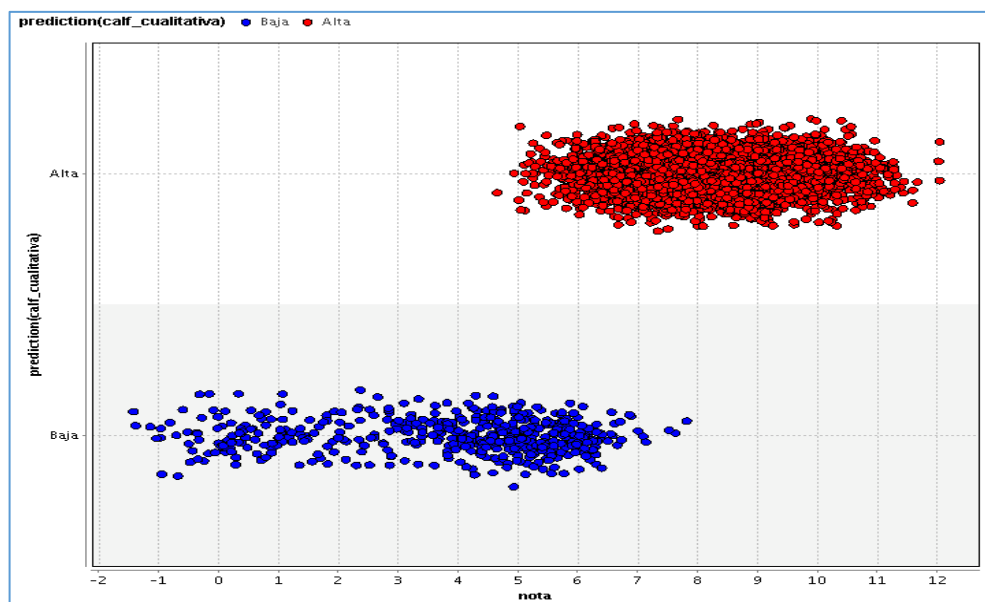


Figura 50: Descripción y análisis gráfico.

Reforzando lo antes mencionado se puede observar que a mayor calificación de un estudiante mayor también será la calificación hacia un docente.

### 3.1.4.3.2 Clasificación mediante ID3

Mediante el árbol de clasificación se pudo observar de manera gráfica y estructurada el comportamiento de los estudiantes al momento de evaluar a un docente. Los parámetros que se establecieron fueron el número mínimo de divisiones que se pueden dar por cada nodo (*minimal size for split=3*), el tamaño mínimo de cada hoja (*minimal leaf size=2*) y la ganancia mínima (*minimal gain=0.6*), (*information\_gain*) como ganancia de información, en la validación cruzada se seleccionó 7 subconjuntos para evaluar el modelo generado.

Los resultados dados por el algoritmo ID3 son los siguientes (Ver Tabla XCVIII):

TABLA XCVIII: RENDIMIENTO, MODELO, ALGORITMO ID3

<i>Algoritmo ID3</i>		
	Descripción	
	Desempeño Modelo	Validación Cruzada
Valores correctamente clasificados	88.96%	88.50%
Valores incorrectos en la clasificación	11.04%	11.50%

A continuación se presenta la matriz de confusión para el presente modelo (Ver Figura 51):

accuracy: 88.50% +/- 0.50% (mikro: 88.50%)			
	true Baja	true Alta	class precision
pred. Baja	1081	253	81.03%
pred. Alta	1070	9102	89.48%
class recall	50.26%	97.30%	

Figura 51: Matriz de confusión algoritmo ID3

La matriz refleja que mediante el algoritmo ID3 la precisión en la predicción en cada una de las categorías de calificación es de 81.03% y un 89.48% respectivamente. En el siguiente árbol se puede observar la clasificación de los datos por ganancia de información (Ver Figura 52).

```

Tree

periodo_academico = 1
|  nota = buena
|  |  carrera = Ing.Elect y Telec
|  |  |  semestre = 10: Alta {Baja=0, Alta=19}
|  |  |  semestre = 2: Alta {Baja=12, Alta=61}
|  |  |  semestre = 4: Alta {Baja=3, Alta=76}
|  |  |  semestre = 6: Alta {Baja=5, Alta=59}
|  |  |  semestre = 8: Alta {Baja=15, Alta=69}
|  |  carrera = Ing.Electm
|  |  |  semestre = 10: Alta {Baja=9, Alta=101}
|  |  |  semestre = 2: Alta {Baja=12, Alta=100}
|  |  |  semestre = 4: Alta {Baja=8, Alta=65}
|  |  |  semestre = 6: Alta {Baja=5, Alta=25}
|  |  |  semestre = 8: Alta {Baja=4, Alta=27}
|  |  carrera = Ing.Geo y OT
|  |  |  semestre = 10: Alta {Baja=0, Alta=10}
|  |  |  semestre = 2: Alta {Baja=4, Alta=77}
|  |  |  semestre = 4: Alta {Baja=13, Alta=61}
|  |  |  semestre = 6: Alta {Baja=3, Alta=95}
|  |  |  semestre = 8: Alta {Baja=6, Alta=40}

```

Figura 52: Árbol de clasificación ID3

El árbol clasifica el comportamiento de las calificaciones docentes en base a las notas por asignatura de un estudiante y calificación individual hacia un docente. La clasificación mediante el algoritmo ID3 también evidencia que los módulos que más bajo califican a un docente cuando los estudiantes poseen una nota mala, son los primeros módulos.

Para obtener una mejor perspectiva de lo que sucede en la clasificación se discretiza las notas de los estudiantes. El árbol describe que a mayor nota de un estudiante mayor calificación de un docente.

### 3.1.4.3.3 Clasificación mediante C4.5.

Se generó un modelo con el árbol de decisión, el cual permite la descripción de los atributos mediante una clasificación sucesiva de los mismos. El objetivo es crear un modelo de clasificación que predice el valor de un atributo de destino (a menudo llamado clase o etiqueta) basado en varios atributos de entrada. Cada nodo interior del árbol corresponde a uno de los atributos de entrada. Cada nodo hoja representa un valor del atributo de la etiqueta dados los valores de los atributos de entrada representados por el camino desde la raíz a la hoja. Esta descripción puede ser fácilmente entendida de fácil interpretación y análisis.

**Para el mismo se fijó los siguientes parámetros:**

La longitud mínima para el parámetro de división en cada nodo, el subconjunto o tamaño mínimo de un nodo hoja, la ganancia mínima de un nodo antes de su división, la profundidad máxima o tamaño del árbol de decisión, el nivel de confianza utilizado para el cálculo de error.

Al desactivar las opciones de *no prepruning* y *preuning*, se obtiene un árbol sin una pre poda y al finalizar el modelo un árbol sin podar para poder obtener una clasificación óptima.

- ✓ minimal size for split= 3
- ✓ minimal leaf size= 2
- ✓ minimal gain=0.6
- ✓ *maximal deph*=10
- ✓ *confidence*=0.5
- ✓ *number of prepruning*=3

Los resultados en el modelo son los siguientes (Ver Tabla XCIX):

TABLA XCIX: RENDIMIENTO, MODELO, C4.5

<b>Algoritmo C4.5</b>		
	<b>Descripción</b>	
	<b>Desempeño Modelo</b>	<b>Validación Cruzada</b>
Valores correctamente clasificados	88.68%	88.73%
Valores incorrectos en la clasificación	11.32%	11.27%

La matriz de confusión en la validación cruzada arrojó los siguientes resultados (Ver Figura 53):

accuracy: 88.73% +/- 0.80% (mikro: 88.73%)			
	true Baja	true Alta	class precision
pred. Baja	1096	242	81.91%
pred. Alta	1055	9113	89.62%
class recall	50.95%	97.41%	

Figura 53: Matriz de confusión C4.5

La matriz de confusión revela que existe una alta clasificación para las categorías Baja 81.91% y Alta con un 89.62%. A continuación se presenta un estimado del árbol de decisión generado (ver figura 54):

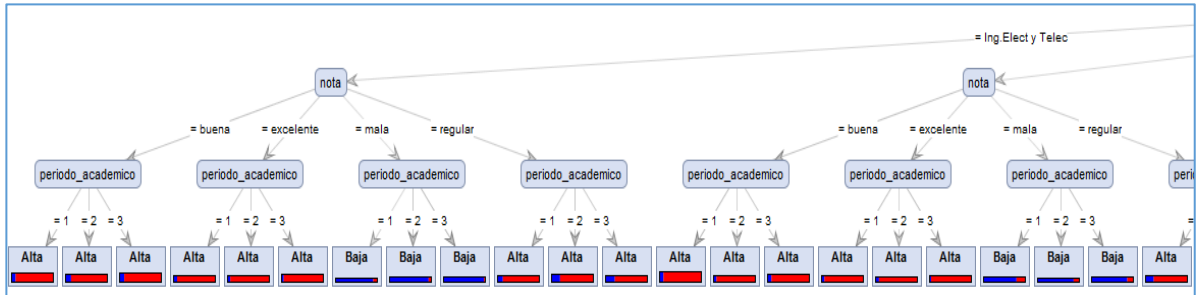


Figura 54: Árbol de decisión generado C4.5

La figura muestra una clasificación evidente en cuanto a la influencia de las notas de los estudiantes hacia la calificación docente, es decir mientras más baja la nota de un estudiante más baja la calificación en evaluación de un docente.

## 4. FASE 4: Evaluación del modelo generado e interpretación de Resultados

### 4.1 Evaluación Modelo

Para la evaluación de los modelos se hizo un análisis comparativo que permita determinar el grado de acierto para las predicciones y clasificaciones realizadas, se tomaron como medidas de evaluación el error absoluto y el error relativo, kappa y error cuadrático medio medidas que especifican:

**error absoluto:** es la medida que indica la distancia de acierto entre el valor estimado y el valor real.

**error relativo:** es el cociente entre el valor absoluto y el valor real expresado en porcentaje.

**error cuadrático medio:** mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. El *ecm* es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática.

**kappa:** es una medida de coherencia entre las clases pronosticadas por el clasificador y las clases observadas, que tiene en cuenta las posibles correspondencias debidas al azar. Si el valor se aproxima a 1 existe una coherencia perfecta, o si es 0 es una coherencia debida al azar, pero si el valor es negativo entonces es una coherencia menor que la que se espera por azar.

#### 4.1.1 Criterios de Interés existentes en el proceso de evaluación docente.

Para la evaluación de los modelos se hizo un análisis comparativo que permita determinar el grado de concordancia o la confiabilidad en cada uno de los algoritmos aplicados. A continuación se presenta el rendimiento de cada algoritmo y modelo aplicado (Ver Tabla C).

TABLA C: EVALUACIÓN DE MODELOS GENERADOS EN CRITERIOS DE INTERÉS.

Modelo	Evaluación	Valores correctamente clasificados	Valores incorrectos en la clasificación	Error absoluto	Error Relativo	Kappa	Error cuadrático medio
ID3	Desempeño Modelo	100%	0.00%	0.000	0.00%	1.000	0.000
	Validación Cruzada	97.25%	2.75%	0.026	2.58%	0.803	0.156
	Desempeño	99.82%	0.18%	0.003	0.34%	0.987	0.041



<b>CHAID</b>	Modelo						
	Validación Cruzada	98.90%	1.10%	0.011	1.13%	0.924	0.046
<b>C4.5</b>	Desempeño Modelo	96.70%	3.30%	0.061	6.05%	0.719	0.174
	Validación Cruzada	94.32%	5.68%	0.084	8.41%	0.500	0.222
<b>NNge</b>	Desempeño Modelo	100%	0.00%	0.000	0.00%	1.000	0.00
	Validación Cruzada	98.89%	1.10%	0.011	7.65%	0.924	0.105
<b>PART</b>	Desempeño Modelo	99.82%	0.18%	0.003	0.29%	0.987	0.038
	Validación Cruzada	99.27%	0.73%	0.011	1.10%	0.944	0.064

En la tabla se puede observar que existe un mínimo de error de clasificación en cada uno de los algoritmos, en el desempeño del modelo se puede observar que existe un elevado porcentaje de clasificación debido a que los datos que sirvieron en la generación de las reglas, dando así una sobreestimación de los resultados, pero luego con la aplicación de validación cruzada se obtuvieron resultados más reales y exactos, es por ello que algunos de estos algoritmos bajaron su porcentaje de clasificación como son los algoritmos ID3, C4.5 Y NNge, la siguiente figura presenta la evaluación de los algoritmos aplicados. (Ver Figura 55).

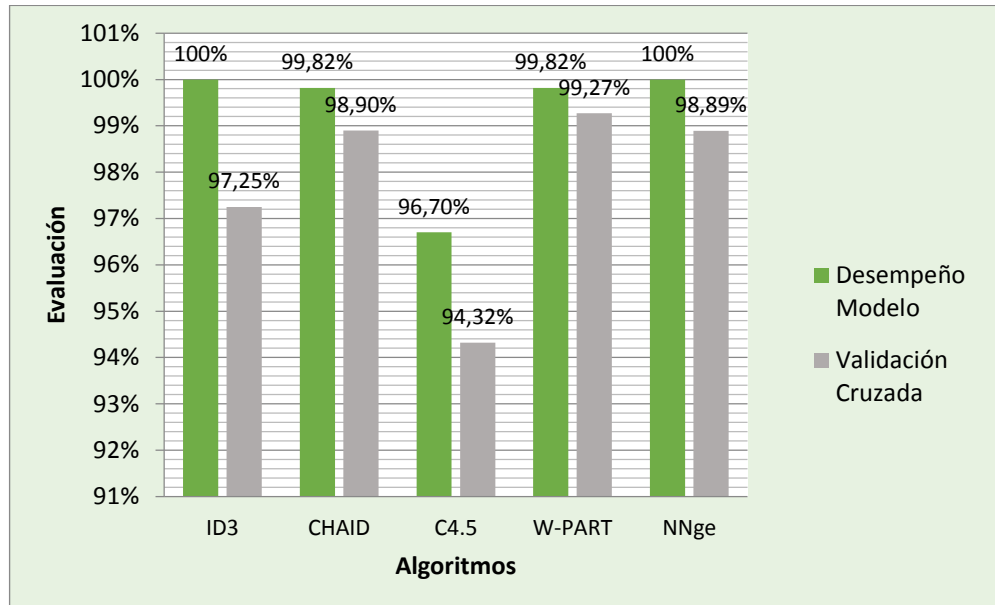


Figura 55: Resultados obtenidos por cada algoritmo Criterios Interés.

Se puede concluir que el algoritmo que mejor se desempeña con los datos disponibles es el algoritmo ID3, CHAID, NNge y W-PART. En conclusión los algoritmos aplicados obtuvieron buenos resultados y útiles dentro del análisis propuesto, descartando los 2 últimos debido a la poca generación de reglas [110].

Los algoritmos con mejor rendimiento en cuanto a las medidas de error obtenidas, sin tomar en cuenta los descartados son los siguientes:

**Error absoluto:** CHAID, con 0.011

**Error relativo:** CHAID con 1.13%

**Kappa:** CHAID, ID3 con 0.924 y 0.803

**Error cuadrático medio:** CHAID, ID3 con 0.046 y 0.156

En la siguiente figura se describen los porcentajes tanto para criterios de interés como para buen rendimiento (ver figura 56).

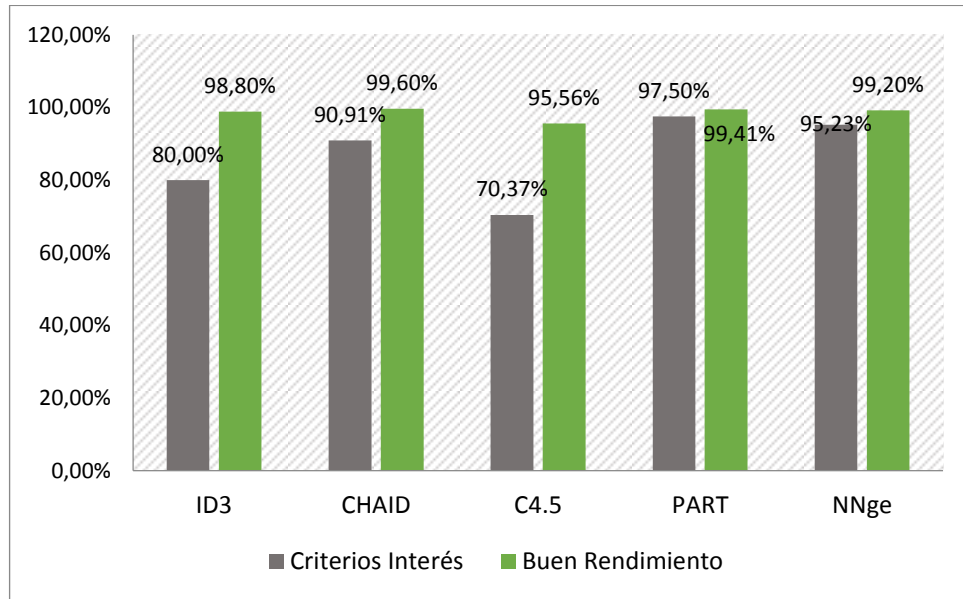


Figura 56: Resultados obtenidos por cada clase Criterios Interés.

La imagen revela que los algoritmos que mejor resultado lograron, en cuanto a la clase criterios de interés son: ID3, CHAID, PART, NNge, como se mencionó anteriormente los algoritmos PART y NNge aunque tengan buena clasificación son descartados debido a que los mismos generan pocas y redundantes reglas [110].

Por lo tanto se decidió trabajar con los algoritmos ID3 y CHAID, ya que los mismos presentan una buena clasificación obtenida de validación cruzada que sobre pasa al 90% de clasificación y una baja en cuanto a las medidas de error.

#### 4.1.2 Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos.

Para la evaluación de los modelos se hizo un análisis comparativo que permita determinar el grado de concordancia o la confiabilidad en cada uno de los algoritmos aplicados.

A continuación se presenta el rendimiento de cada algoritmo y modelo aplicado (Ver Tabla CI).

TABLA CI: EVALUACIÓN DE MODELOS GENERADOS EVOLUCIÓN DE CÁTEDRA.

Modelo	Evaluación	Valores correctamente clasificados	Valores incorrectos en la clasificación	Error absoluto	Error Relativo	Kappa	Raíz error cuadrático medio
Naive Bayes	Desempeño Modelo	93.05%	6.95%	0.076	7.64 %	0.844	0.256
	Validación Cruzada	93.05%	6.95%	0.078	7.84%	0.845	0.255
ID3	Desempeño Modelo	97.30%	2.70%	0.035	3.52%	0.939	0.133
	Validación Cruzada	92.28%	7.72%	0.079	7.86%	0.820	0.241
REPTree	Desempeño Modelo	96.14%	3.86%	0.050	5.03%	0.911	0.159
	Validación Cruzada	94.21%	5.79%	0.077	7.66%	0.858	0.212
NNge	Desempeño Modelo	96.91%	3.08%	0.031	7.04%	0.929	0.175
	Validación Cruzada	94.21%	5.79%	0.057	13.19%	0.867	0.240

En la tabla se puede observar que existe un alto grado de concordancia en todos los modelos generados por los algoritmos.

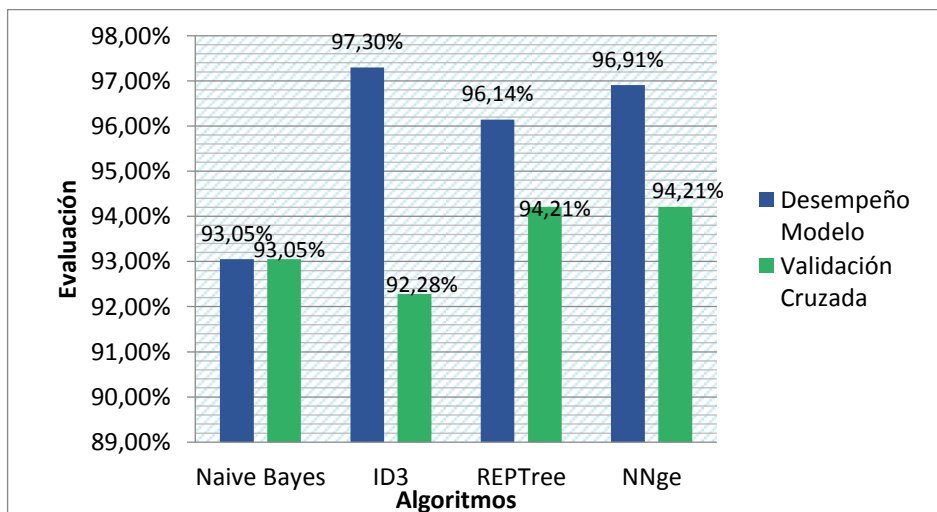


Figura 57: Resultados obtenidos por cada algoritmo- Evolución Cátedra.

Se puede concluir que el algoritmo que mejor se desempeña con los datos disponibles es el algoritmo REPTree, NNge y Naive Bayes, sin descartar los resultados arrojados por el algoritmo ID3 ya que tiene una precisión mayor del 90%. En conclusión los algoritmos aplicados obtuvieron buenos resultados y útiles dentro del análisis propuesto, descartando los REPTree debido a que poseía reglas difíciles de interpretar [110].

Los algoritmos con mejor rendimiento en cuanto a las medidas de error obtenidas, sin tomar en cuenta el descartado son los siguientes:

**Error absoluto:** NNge con 0.057

**Error relativo:** Naive Bayes con 7.84%

**Kappa:** NNge con 0.867

**Error cuadrático medio:** NNge, ID3 con 0.240 y 0.241

En la siguiente figura se describen los porcentajes tanto para evolución de cátedra (Si) como para la ausencia de la misma (No), (ver figura 58).

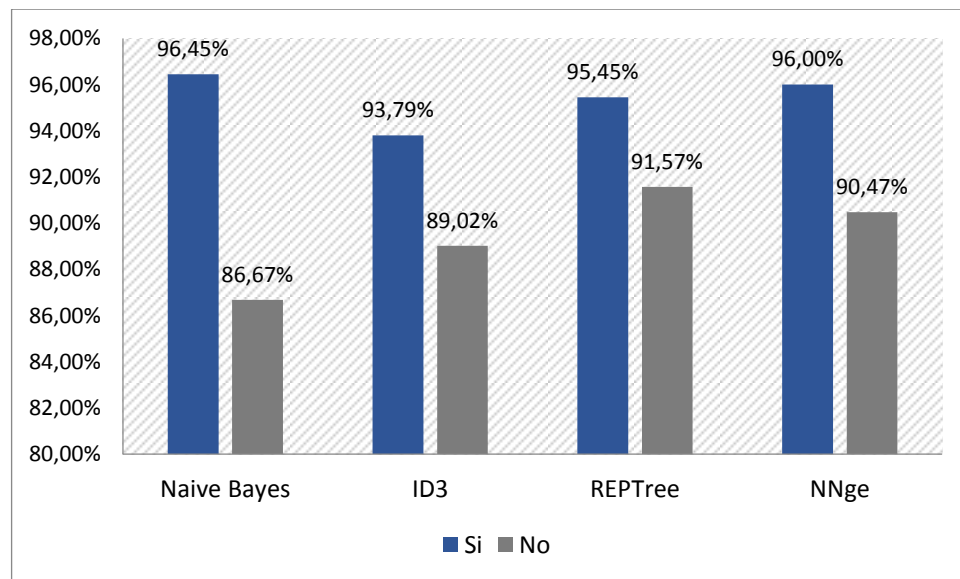


Figura 58: Resultados obtenidos por cada clase, Evolución Cátedra

La imagen revela que los algoritmos que mejor resultado lograron, en cuanto a la clase Evolución de cátedra (Si) son: Naive Bayes, NNge y REPTree, como se mencionó anteriormente el algoritmo REPTree aunque tengan buena clasificación es descartado debido a que genera reglas redundantes [110].

Por lo tanto se decidió trabajar con los algoritmos Naive Bayes, NNge e ID3 ya que los mismos presentan una buena clasificación obtenida de validación cruzada que sobre pasa al 90% de clasificación y una baja en cuanto a las medidas de error.

#### 4.1.3 Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.

Para la evaluación de los modelos se hizo un análisis comparativo que permita determinar el grado de concordancia o la confiabilidad en cada uno de los algoritmos aplicados.

A continuación se presenta el rendimiento de cada algoritmo y modelo aplicado (Ver Tabla CII).

TABLA CII: EVALUACIÓN DE MODELOS GENERADOS EN INFLUENCIA DE NOTAS.

Modelo	Evaluación	Valores correctamente clasificados	Valores incorrectos en la clasificación	Error absoluto	Error Relativo	kappa	Raíz error cuadrático medio
<i>Regresión Logística</i>	<b>Desempeño Modelo</b>	88.77%	11.23%	0.206	20.59%	0.536	0.315
	<b>Validación Cruzada</b>	88.74 %	11.26 %	0.207	20.69 %	0.528	0.316
Árbol de Clasificación ID3	<b>Desempeño Modelo</b>	88.96 %	11.04 %	0.179	17.86 %	0.579	0.298
	<b>Validación Cruzada</b>	88.50 %	11.50 %	0.187	18.66 %	0.557	0.309
<i>Árbol de decisión C4.5</i>	<b>Desempeño Modelo</b>	88.68 %	11.32 %	0.191	19.07 %	0.558	0.308
	<b>Validación Cruzada</b>	88.73%	11.27 %	0.192	19.23 %	0.566	0.310

En la tabla se puede observar que existe un alto grado de precisión en los modelos de regresión logística y el árbol de clasificación C4.5, a diferencia del árbol de decisión ID3 si bien es cierto baja el grado de precisión, sus resultados y condiciones pueden ser tomadas en cuenta en base a los modelos anteriores.

Cabe destacar que los algoritmos se desempeñan de acuerdo a los datos disponibles, por lo general al aplicar un modelo a un conjunto de datos de prueba el rendimiento aumenta

y cuando se aplica una validación cruzada el desempeño varía, en unas ocasiones baja y en otras ocasiones el algoritmo se desempeña mejor a mayor información obtenida. A diferencia de los modelos anteriores, se puede afirmar una vez más que la calidad de los datos influye en la generación de los modelos en este caso la existencia de valores nulos [109], (Ver Figura 59).

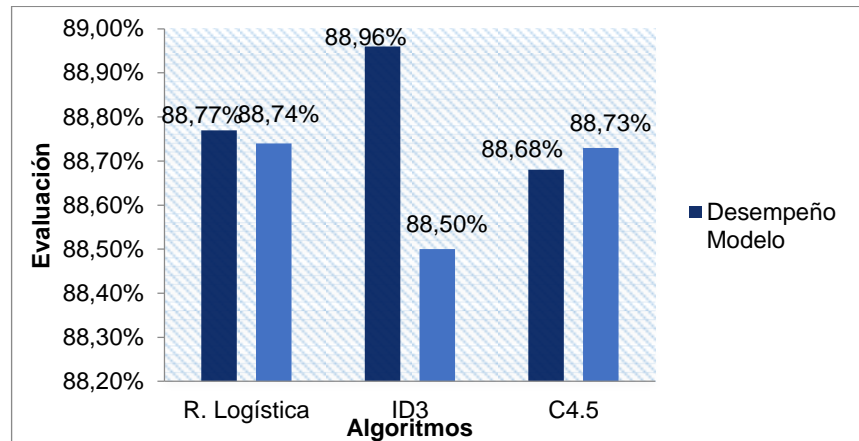


Figura 59: Resultados obtenidos por cada algoritmo, Notas.

Se puede concluir que los algoritmos que mejor se desempeñan con los datos disponibles el algoritmo de Regresión Logística y el algoritmo ID3, sin embargo existen condiciones en el árbol de decisión C4.5 que pueden ser tomados en cuenta dentro del análisis.

Los algoritmos con mejor rendimiento en cuanto a las medidas de error obtenidas, son los siguientes:

**Error absoluto:** ID3 con 0.187

**Error relativo:** ID3 con 18.66 %

**Kappa:** ID3 con 0.557

**Error cuadrático medio:** ID3 con 0.309

En la siguiente figura se describen los porcentajes tanto para la clase Alta y Baja (ver figura 60).

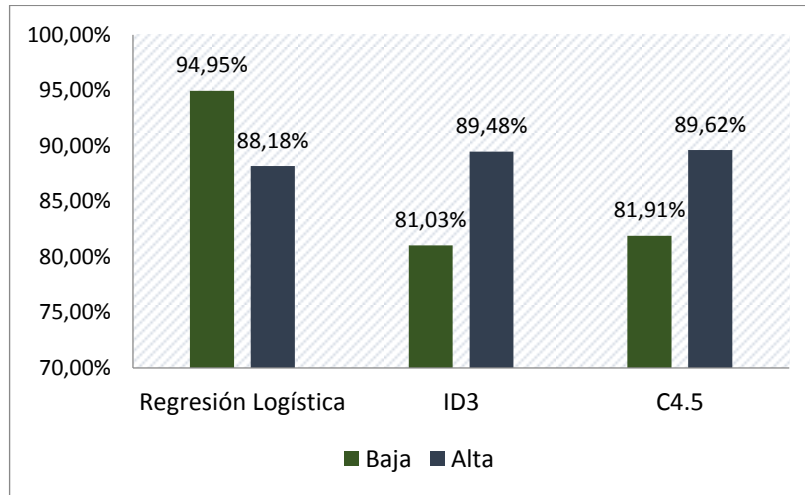


Figura 60: Resultados obtenidos por cada clase, Notas.

La imagen revela que los algoritmos que mejor resultado lograron, en cuanto a la clase Baja es ID3, como se mencionó anteriormente, se puede afirmar una vez más que la calidad de los datos influye en la generación de los modelos en este caso la existencia de valores nulos [109],

Por lo tanto se decidió trabajar con los algoritmos propuestos: Regresión Logística, ID3 y C4.5 ya que los mismos presentan una buena clasificación obtenida de validación cruzada que sobre pasa al 87% de clasificación debido a la calidad de los datos [111].

## 4.2 Validación del Modelo y Análisis de Resultados.

Luego de evaluar los algoritmos y modelos generados se realizó la validación del modelo y análisis de resultados presentados a continuación.

### 4.2.1 Evaluación y Validación

En la presente fase se recopiló la información, comportamientos o factores relevantes generados por los modelos diseñados. Si bien es cierto algunos modelos al momento de ser evaluados bajaron su rendimiento en comparación con otros, estos pueden dar una pauta de análisis o descripción de características no tan evidentes en otros modelos.



### 4.2.1.1 Evaluación de Resultados

Los modelos que presentaron un mejor rendimiento se les dio mayor énfasis al presentar los resultados, no dejando de lado a su vez, los modelos que presentaron un menor rendimiento esto con el fin de visualizar y analizar toda la información generada y en diferentes perspectivas (Ver Anexo 8).

#### 4.2.1.1.1 Resultados de Criterios de Interés existentes en el proceso de evaluación docente.

En este punto primero se obtuvo el modelo para poder identificar donde y cuando existen docentes con criterios de interés (Ver Figura 61), para luego aplicar minería de textos sobre las opiniones de los estudiantes a través de la herramienta Many Eyes, así mismo con el buen rendimiento que presentan los docentes.

Se entiende por criterios de interés a las opiniones vertidas por los estudiantes dentro del proceso de evaluación, las cuales pueden revelar circunstancias o comportamientos que un docente debe reforzar, además cuando los docentes no posean criterios de interés, se presenta las opiniones que hacen referencia al buen rendimiento o nivel de satisfacción que puede tener un estudiante con respecto a la forma de cátedra de un docente, en una carrera en un periodo académico y una asignatura determinada.

El buen rendimiento de igual manera hace referencia a las opiniones de estudiantes, pero en este caso especifican las fortalezas que tienen los docentes en un periodo de evaluación, las mismas que pueden ser tomadas en cuenta en la toma de decisiones.

- **Modelo ID3**

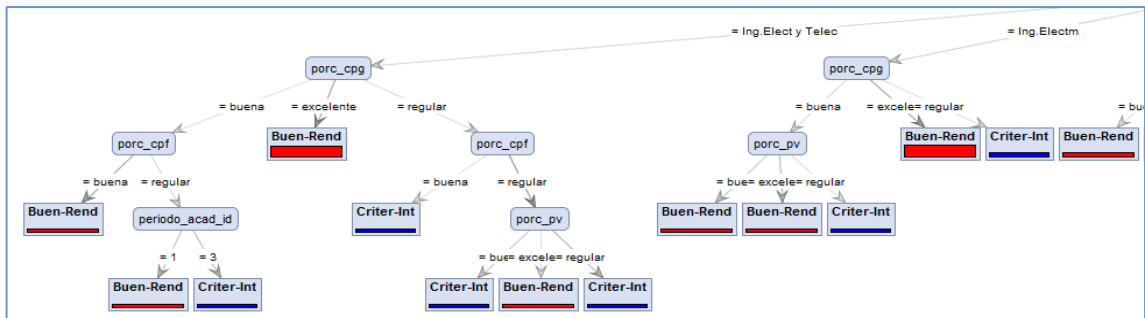


Figura 61: Criterios de interés y Buen rendimiento mediante algoritmo ID3.

El árbol muestra las instancias clasificadas en la existencia de criterios de interés y buen rendimiento tomando como raíz cada una de las carreras del área de la energía.

El árbol evidencia que el componente de calificación más importante dentro de la evaluación es el componente pedagógico, es decir si un docente posee un buen porcentaje de calificación en su capacidad pedagógica entonces tendrá un buen rendimiento dentro de la evaluación de desempeño, pero cabe excepciones que se especifican más adelante. La carrera de Tecnología en Electricidad y Control Industrial en el análisis no tiene criterios de interés en ningún periodo académico, teniendo solo docentes con un buen rendimiento en todos los periodos de evaluación, la siguiente figura (Ver Figura 62) presenta de manera gráfica lo descrito por el árbol de clasificación ID3.

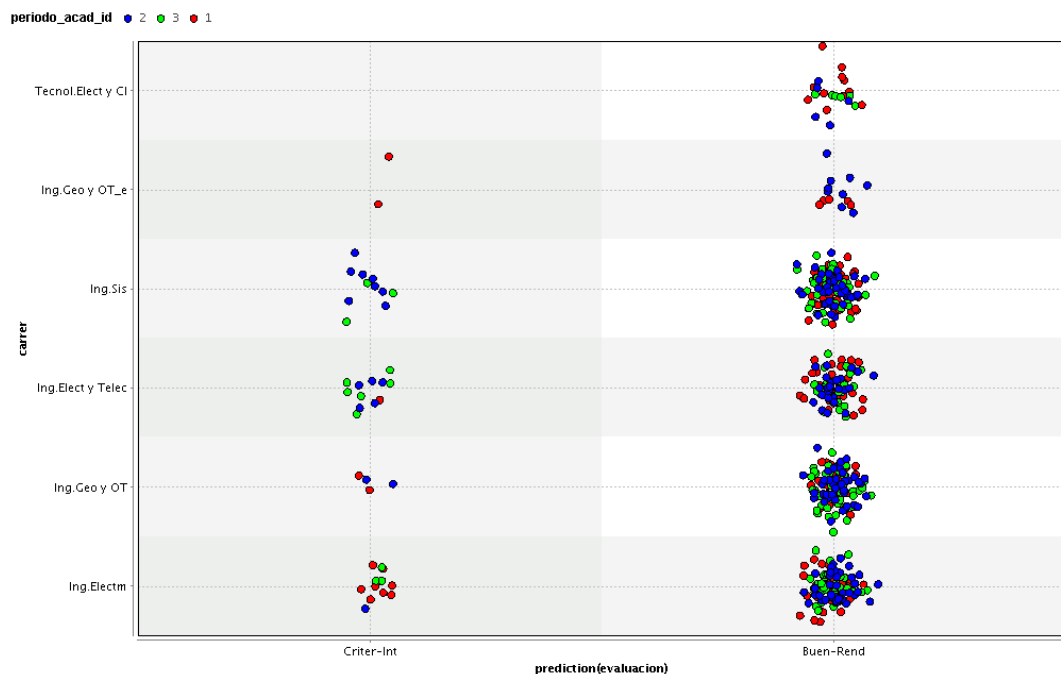


Figura 62: Descripción y análisis gráfico.

En base a la imagen se puede describir que todas las carreras poseen un buen rendimiento. Además se puede evidenciar que en el periodo académico 2 se da la mayor existencia de criterios de interés.

La estructura sobre la cual se llevó a cabo la ejecución del modelo para obtener los criterios de interés, tiene un registro de 545 docentes que pertenecen a todas las carreras,

los mismos que han dictado una o varias materias en los periodos de evaluación, en la figura 63 se presentan las instancias clasificadas del modelo generado.

Role	Name	Type	Sum	Missings
id	cedula_docent	nominal	?	0
label	evaluacion	nominal	?	0
confidence_Criter-Int	confidence(Criter-Int)	real	42	0
confidence_Buen-Rend	confidence(Buen-Rend)	real	503	0

Figura 63: Docentes clasificados mediante el algoritmo ID3.

En la figura anterior se puede observar que existen 503 docentes con un buen rendimiento y 42 que poseen criterios de interés, y que equivale al 92.29%, 7.71% respectivamente.

- **CHAID**

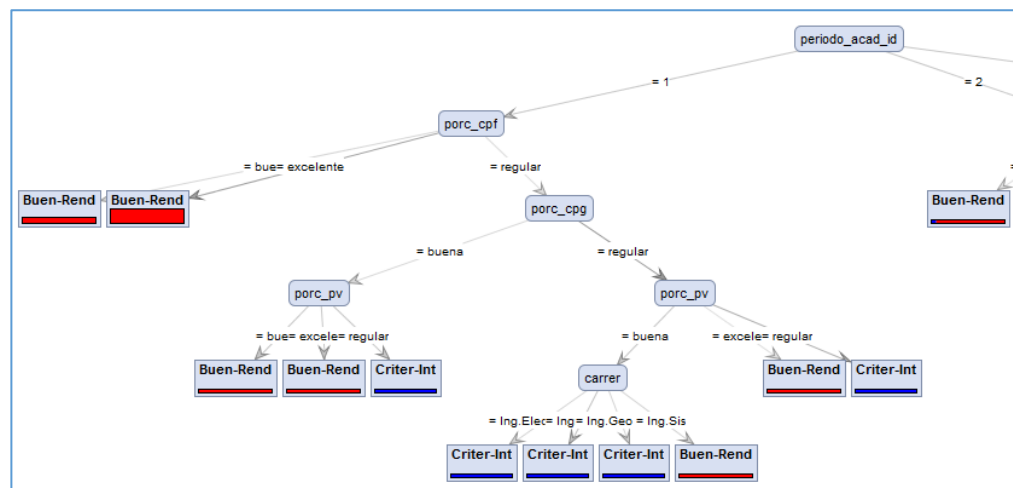


Figura 64: Criterios de interés y Buen rendimiento mediante algoritmo CHAID.

El árbol muestra las instancias clasificadas en la existencia de criterios de interés y buen rendimiento toman como raíz los periodos de evaluación.

En este algoritmo se desglosa los datos desde el porcentaje de calificación en capacidad profesional hacia la capacidad pedagógica y la práctica de valores, que se le otorga a un docente, dando la clasificación en la existencia de criterios de interés o buen rendimiento según las condiciones especificadas por el árbol generado, la siguiente figura (Ver Figura 65) presenta las instancias clasificadas del modelo generado.

Role	Name	Type	Sum	Missings
id	cedula_docent	nominal	?	0
label	evaluacion	nominal	?	0
confidence_Criter-Int	confidence(Criter-Int)	real	42	0
confidence_Buen-Rend	confidence(Buen-Rend)	real	503	0

Figura 65: Docentes clasificados mediante el algoritmo CHAID.

En la figura anterior se puede observar que existen 503 docentes con un buen rendimiento y 42 que poseen criterios de interés, y que equivale al 92.29%, 7.71% respectivamente.

En base a la imagen se puede describir que todas las carreras poseen un buen rendimiento. Además nuevamente se evidencia que en el periodo académico 2 se da la mayor existencia de criterios de interés y el primer periodo académico es donde existen más docentes con un buen rendimiento en las asignaturas dictadas.

A diferencia del modelo anterior en la aplicación del presente algoritmo se ha tomado cada una de las asignaturas en donde se da existencia o ausencia de criterios de interés en base a las asignaturas dictadas dentro del área de la energía, de las cuales se puede rescatar las siguientes (Ver Tabla CIII)

TABLA CIII: ASIGNATURAS QUE GENERAN CRITERIOS DE INTERÉS.

Carrera	Asignaturas
Ing. Electromecánica	<ul style="list-style-type: none"> <li>▪ análisis matemático</li> <li>▪ teoría electromagnética</li> <li>▪ accionamiento eléctrico</li> <li>▪ estática</li> <li>▪ maquinas eléctricas</li> <li>▪ proyecto electro energético</li> <li>▪ proyecto de tesis</li> <li>▪ control de la producción y calidad</li> <li>▪ sistemas eléctricos</li> <li>▪ trigonometría</li> <li>▪ instalaciones eléctricas</li> <li>▪ Sistemas eléctricos</li> </ul>
	<ul style="list-style-type: none"> <li>▪ electrónica de potencia</li> </ul>

<p>Ing. Electrónica y Telecomunicaciones</p>	<ul style="list-style-type: none"> <li>▪ prácticas de circuitos electrónicos de potencia</li> <li>▪ prácticas de comunicaciones digitales</li> <li>▪ análisis de señales y sistemas</li> <li>▪ redes alámbricas de última generación</li> <li>▪ sistemas de comunicaciones satelitales</li> <li>▪ televisión convencional</li> <li>▪ formulación y evaluación de proyectos</li> <li>▪ circuitos eléctricos</li> <li>▪ prácticas de circuitos eléctricos</li> <li>▪ proyecto de ingeniería</li> <li>▪ física</li> </ul>
<p>Ing. Geología Ambiental y Ordenamiento Territorial</p>	<ul style="list-style-type: none"> <li>▪ biología</li> <li>▪ algebra lineal</li> <li>▪ investigación</li> <li>▪ prevención de riesgos</li> <li>▪ geodinámica</li> <li>▪ prevención contaminación</li> </ul>
<p>Ing. Sistemas</p>	<ul style="list-style-type: none"> <li>▪ proceso investigativo</li> <li>▪ teoría de telecomunicaciones</li> <li>▪ derecho informático</li> <li>▪ estadística inferencial</li> <li>▪ arquitectura de computadores</li> <li>▪ ingeniería del software</li> <li>▪ programación avanzada</li> <li>▪ matemáticas discretas</li> <li>▪ algebra lineal</li> <li>▪ física</li> <li>▪ teoría de la programación</li> </ul>

Por motivo de confidencialidad se descarta la cédula de los docentes dentro del análisis y se toma las asignaturas en donde existen criterios de interés, de igual manera se toman las asignaturas más relevantes, es decir las que han tenido mejor rendimiento dentro de los periodos de evaluación en cada carrera del área de la energía (Ver Tabla CIV)

TABLA CIV: ASIGNATURAS BUEN RENDIMIENTO.

<b>Carrera</b>	<b>Asignaturas</b>
Ing. Electromecánica	<ul style="list-style-type: none"> <li>▪ diseño mecánico</li> <li>▪ materiales para ingeniería</li> <li>▪ energías alternativas</li> <li>▪ matemáticas</li> <li>▪ calculo diferencial</li> <li>▪ economía para ingenieros</li> <li>▪ mantenimiento y seguridad industrial</li> <li>▪ química</li> </ul>
Ing. Electrónica y Telecomunicaciones	<ul style="list-style-type: none"> <li>▪ comunicaciones digitales</li> <li>▪ electrónica básica</li> <li>▪ legislación de las telecomunicaciones</li> <li>▪ telefonía celular</li> <li>▪ comunicaciones analógicas</li> <li>▪ comunicaciones ópticas</li> <li>▪ ingeniería de tráfico</li> <li>▪ prácticas de comunicaciones analógica</li> <li>▪ comunicaciones digitales</li> <li>▪ nanotecnología y nuevos materiales</li> </ul>
Ing. Geología Ambiental y Ordenamiento Territorial	<ul style="list-style-type: none"> <li>▪ investigación del modulo</li> <li>▪ matemática básica</li> <li>▪ mecánica de suelos</li> <li>▪ ordenamiento territorial en obras civiles</li> <li>▪ trigonometría</li> <li>▪ computación</li> <li>▪ elaboración de proyecto</li> <li>▪ estadística</li> <li>▪ investigación</li> <li>▪ matemáticas</li> <li>▪ obras y construcciones civiles</li> <li>▪ análisis matemático</li> <li>▪ física</li> <li>▪ mecánica de suelos</li> </ul>

	<ul style="list-style-type: none"> <li>▪ ordenamiento territorial en obras civiles</li> <li>▪ proceso investigativo</li> </ul>
Ing. Sistemas	<ul style="list-style-type: none"> <li>▪ inteligencia artificial</li> <li>▪ compiladores</li> <li>▪ lenguajes formales</li> <li>▪ sistemas de información</li> <li>▪ anteproyectos de tesis</li> <li>▪ estructura de datos orientada oo</li> <li>▪ administración de centros de computo</li> <li>▪ sistemas operativos</li> </ul>
Tecnología en Electricidad y Control Industrial	<ul style="list-style-type: none"> <li>▪ prácticas de reparación y mantenimiento de máquinas eléctricas rotativas</li> <li>▪ prácticas de reparación y mantenimiento de transformadores</li> <li>▪ proyecto de trabajo practico</li> <li>▪ prácticas de redes de distribución</li> <li>▪ principios generales de centrales y subestaciones</li> <li>▪ redes de distribución</li> <li>▪ electrónica industrial y de potencia</li> <li>▪ instrumentación</li> <li>▪ proyecto de trabajo practico</li> </ul>

• **Análisis Criterios Interés**

Los resultados en este aspecto son los siguientes:

- ✓ En la mayoría de las carreras, si el porcentaje de calificación en cuanto al componente de capacidad pedagógica es excelente entonces habrá un buen rendimiento en el proceso de evaluación.
- ✓ La carrera de Ingeniería en Geología Ambiental y Ordenamiento Territorial (especial), en el periodo académico 1, con un porcentaje de capacidad profesional excelente presenta buen rendimiento en los docentes, pero si la capacidad profesional de un docente es regular entonces existen criterios de interés.
- ✓ La carrera de Ing. en Geología Ambiental y Ordenamiento Territorial (especial) en el periodo académico 2, existe buen rendimiento docente.

- ✓ En la carrera de tecnología de Electricidad y Control Industrial existe la mínima cantidad de docentes con criterios de interés.
- ✓ Las carreras con menos docentes con criterios de interés en el proceso de evaluación son las carreras de Ing. en Geología Ambiental y Ordenamiento Territorial, presencial y especial, seguido de la carrera de Tecnología en Electricidad y Control Industrial.
- ✓ El periodo donde existe la mayor cantidad de docentes con criterios de interés dentro de la evaluación docente es el periodo 2 con un 38.09% del total, y el mayor periodo favorable a buen rendimiento docente es el periodo 1 con un 34.99% del total.
- ✓ En el periodo académico 1 existe un buen rendimiento docente debido a la capacidad profesional de cada docente la cual se clasifica en buena y excelente.
- ✓ En el periodo académico 1 la capacidad profesional y capacidad pedagógica es baja o regular pero existe una buena práctica de valores, entonces la carrera de Ingeniería en Sistemas presenta un Buen Rendimiento en sus docentes.
- ✓ En el periodo académico 3, la práctica de valores y capacidad pedagógica obtenida por un docente es buena, pero existe una baja capacidad profesional entonces existen criterios de interés.

- **Opiniones de interés dentro del proceso (Criterios de Interés)**

En este apartado es necesario señalar que actualmente existe un proceso semejante al proceso de minería de Datos, denominado Text Mining o Minería de Textos, el cual permite la aplicación de algoritmos sobre un texto plano, documentos, palabras, libros, etc. Cabe destacar que minería de textos tiene otras etapas para la generación de modelos, etapas que no conciernen al alcance de la presente investigación.

La minería de textos ha sido objeto de estudio en varias investigaciones obteniendo excelentes resultados en la aplicación de la misma [99].

Por lo tanto luego de la identificación de la existencia de criterios de interés o el buen rendimiento de un docente, se aplicó minería de texto a las opiniones de los estudiantes para poder conocer cuáles son las opiniones con mayor relevancia en la evaluación hacia un docente, esto mediante la herramienta Many Eyes que facilita y reduce el trabajo de minería sobre texto (Ver Figura 66, 67).



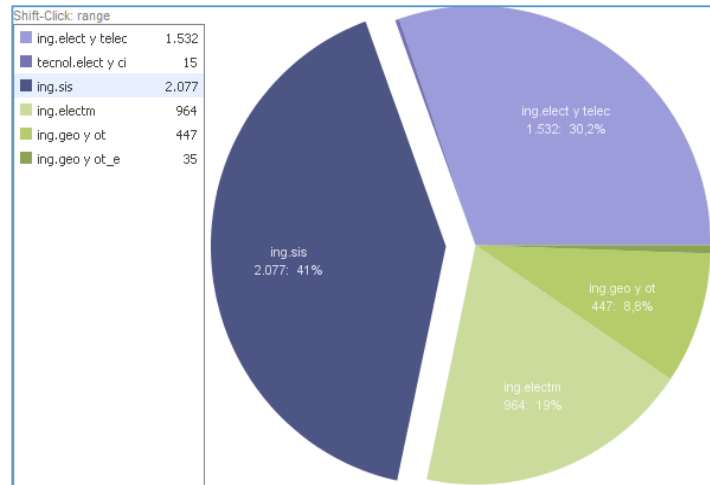


Figura 66: Porcentaje Opiniones Estudiantes por carrera en el proceso de Evaluación

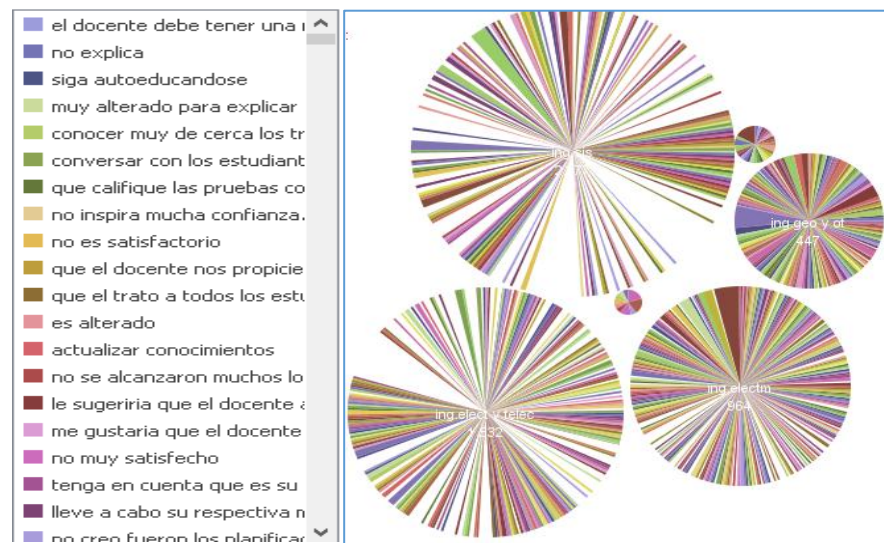


Figura 67: Opiniones Estudiantes dentro del proceso de Evaluación

Las figuras anteriores muestran que la carrera de ingeniería en Sistemas es la carrera con más criterios de interés presenta con un 41% del total de las opiniones dadas por los estudiantes, seguido de la carrera de ingeniería en Electromecánica y Telecomunicaciones con un 30.2% de opiniones, y luego las carreras de ingeniería en Electromecánica, ingeniería en Geología Ambiental y Ordenamiento Territorial con un 19% y 8.8% respectivamente. Además también se puede evidenciar y como se lo había dicho anteriormente que la carrera de Tecnología en Electricidad y control industrial es la carrera con el mínimo de docentes con criterios de interés presenta, es decir las opiniones de los estudiantes en dicha carrera son del 0.3% seguido de la carrera de ingeniería en



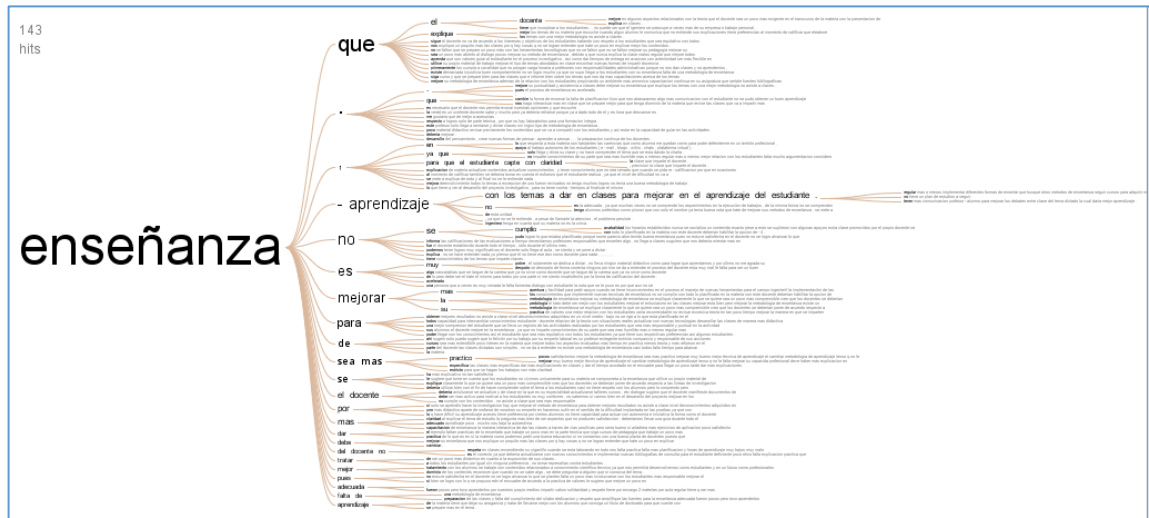


Figura 69: Palabras con mayor ocurrencia: Enseñanza

Dentro de un contexto la palabra enseñanza se repite 143 veces, la misma que hace referencia a varios aspectos en el campo profesional docente, las siguientes palabras son las de mayor ocurrencia, presentadas en orden de mayor incidencia:

- Que
- Aprendizaje
- No

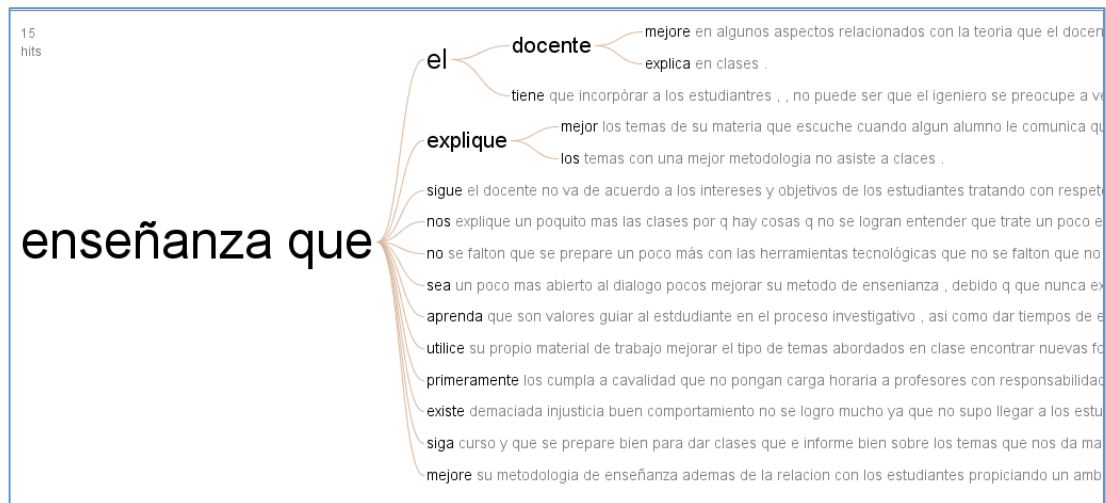


Figura 70: Opiniones Estudiantes dentro del proceso de Evaluación

La palabra enseñanza se enlaza con mayor frecuencia con la palabra (que), por lo tanto se examinan las opiniones dentro de este contexto.

La siguiente figura presenta otra de las palabras con mayor grado de ocurrencias dentro de un contexto, la cual es mejorar.

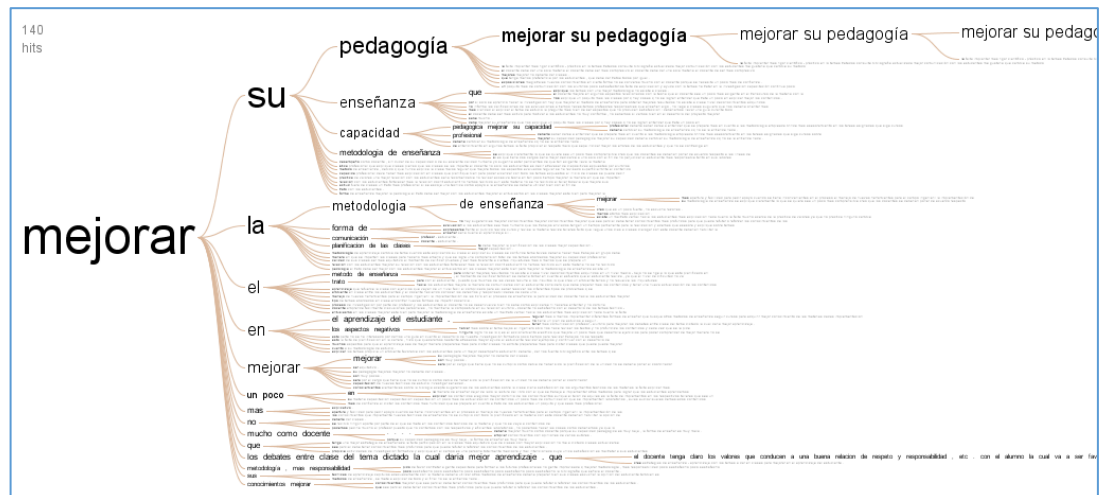


Figura 71: Palabras con mayor ocurrencia: Mejorar

Dentro de un contexto la palabra mejorar se repite 140 veces la misma que hace referencia a mejorar varios aspectos en el campo profesional docente, las siguientes palabras son las de mayor ocurrencia, presentadas en orden de mayor incidencia: Mejorar:

- Su pedagogía
- Su enseñanza
- La metodología de enseñanza.

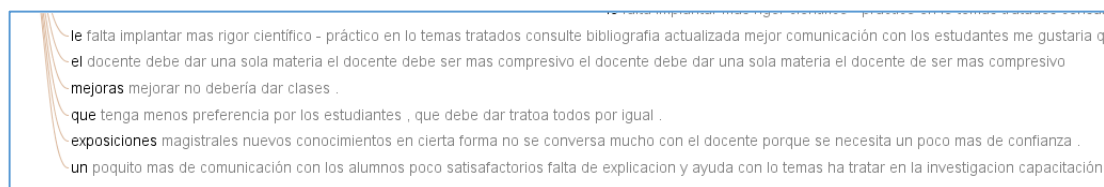


Figura 72: Opiniones Estudiantes dentro del proceso de Evaluación

Existen varias palabras claves, dentro de un plan de mejoras que podrían dar una pauta en la elaboración de las mismas, en base a la opinión del estudiante, por ejemplo (Ver Figura 73)

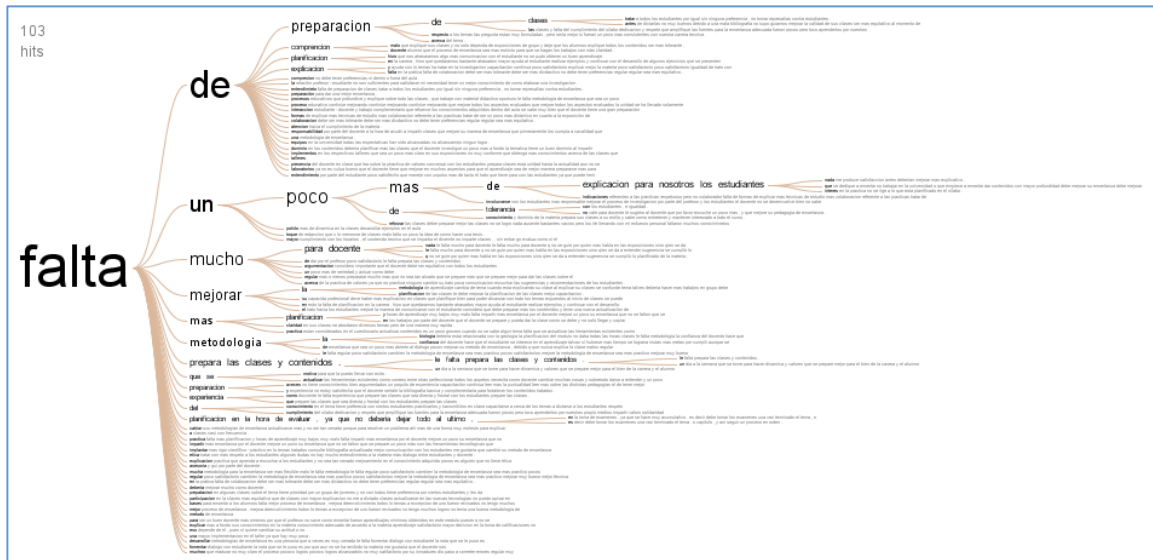


Figura 73: Palabras con mayor ocurrencia: Mejorar

Estas servirían para atacar las falencias existentes en el proceso docente en base a las necesidades evidenciadas (Ver Figura 74)

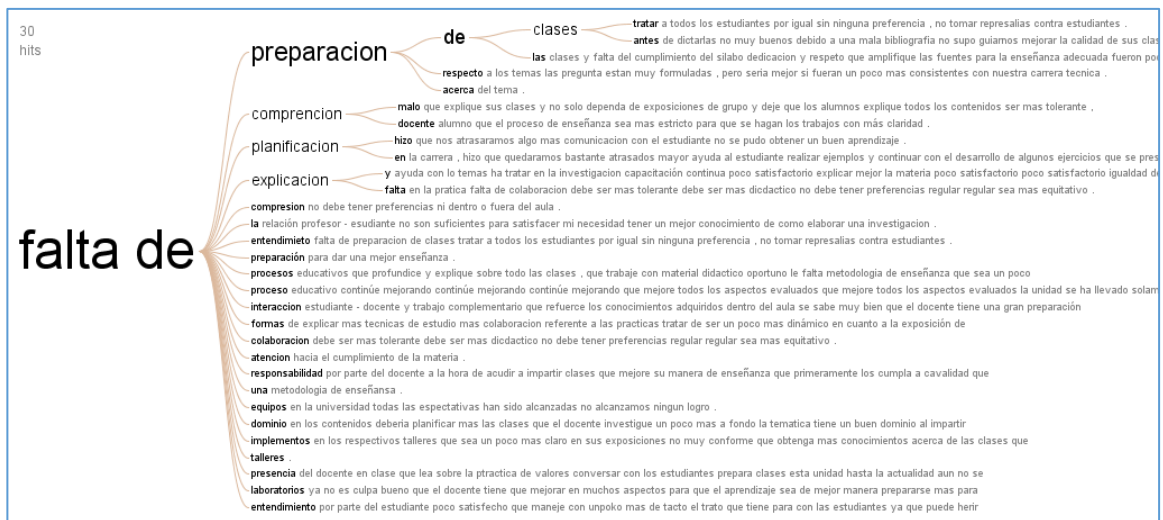


Figura 74: Opiniones Estudiantes: Plan de Mejora

Además se puede reforzar el análisis con el nivel de ocurrencias de las opiniones por carrera la siguiente figura (Ver Figura 75) presenta la cantidad de ocurrencias de una palabra por carrera





En la figura anterior se muestra las opiniones más relevantes en cuanto al buen rendimiento, las cuales son: bien, buen, buenos, buena, excelente, conocimientos, comportamiento, persona, valores, enseñanza.

Todas las palabras obtenidas se encuentran dentro de un contexto, por lo tanto se requiere el análisis de las mismas para saber y observar cómo éstas aportan a un plan de mejoras docentes o ayudan al refuerzo académico. Dicho análisis se presenta en el Anexo 8: Análisis General: Escenario Real.

#### **4.2.1.1.2 Resultados Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos.**

La evolución que tenga un docente en su espacio de cátedra está ligada a su desempeño dentro del aula, a más de las opiniones valorativas que puede dar un estudiante hacia un docente como sugerencia de mejora, también puede reflejarse en el nivel de calificación que ha obtenido un docente en cada materia.

Es decir si un docente ha obtenido una calificación negativa en un periodo, pero en los siguientes tiene una valoración positiva por los estudiantes, evidentemente ha mejorado su metodología de enseñanza, nivel de capacidad pedagógica y por ende la práctica en valores calificados dentro de un cuestionario de desempeño; entonces se puede deducir que el docente ha tenido una evolución de cátedra en sus años de enseñanza.

- **Modelo Naive Bayes**

En el siguiente grafico (Ver Figura 77) se presenta un resumen general de las carreras que han tenido docentes con evolución de cátedra en los 3 periodos académicos:

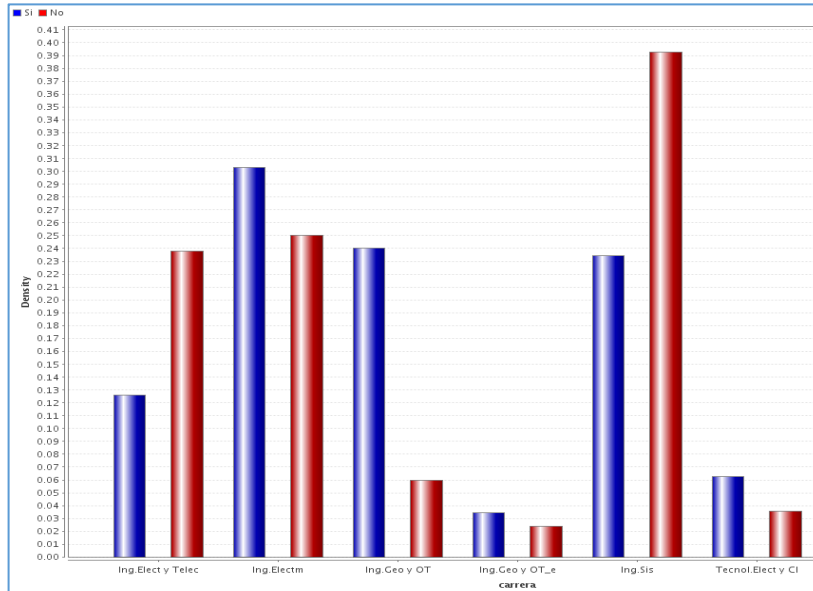


Figura 77: Modelo Naive Bayes

En base a la imagen se puede describir que todas carreras tienen docentes con excelente espacio de cátedra, la carrera que más resalta con una evolución docente es la carrera de Ingeniería en Geología y Ordenamiento Territorial presencial, Ingeniería en Electromecánica, seguido de la carrera de Tecnología en Electricidad y Control Industrial. La siguiente figura muestra la clasificación para cada clase dada por el algoritmo (Ver Figura 78)

Role	Name	Type	Sum	Missings
id	cedula_docent	nominal	?	0
label	evol_catedra	nominal	?	0
confidence_Si	confidence(Si)	real	167.803	0
confidence_No	confidence(No)	real	91.197	0

Figura 78: Docentes clasificados mediante el algoritmo Naive Bayes.

En la figura anterior se puede observar que existen 167 docentes en todas las carreras y todos los periodos académicos con evolución de cátedra y 91 docentes que no tienen una evolución de cátedra en los periodos de estudio correspondiente al 64.47% y 35.13% respectivamente.



- **Modelo ID3**

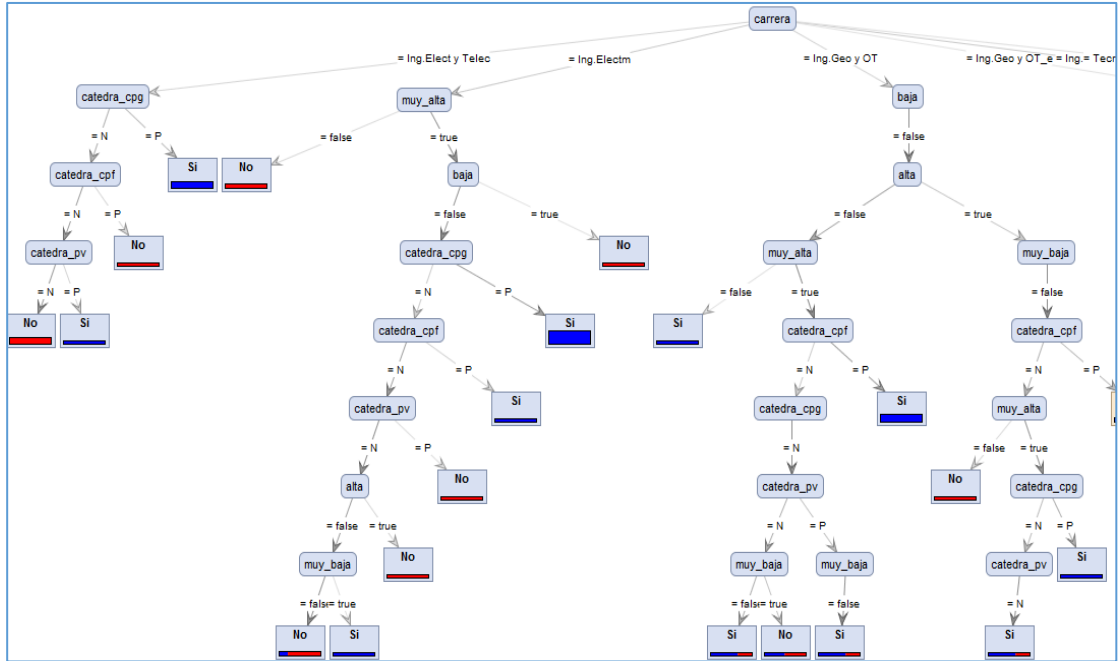


Figura 79: Evolución de cátedra ID3.

El árbol muestra las condiciones para cuando un docente ha tenido o no una evolución de cátedra, se puede observar que en todas las carreras existe evolución de cátedra dadas algunas excepciones.

La estructura sobre la cual se llevó a cabo la ejecución del modelo para obtener la evolución de cátedra obtiene la totalidad de los docentes que han ejercido cátedra dentro del AEIRNNR, en la figura 80 se presentan las instancias clasificadas del modelo generado.

Role	Name	Type	Sum	Missings
id	cedula_docent	nominal	?	0
label	evol_catedra	nominal	?	0
confidence_Si	confidence(Si)	real	175	0
confidence_No	confidence(No)	real	84	0

Figura 80: Docentes clasificados mediante el algoritmo ID3.

En la figura anterior se puede observar 175 que corresponde a un 67.56% de docentes con evolución de cátedra y 84 que corresponde a un 32.43% de docentes sin dicha evolución en los periodos de estudio.

- **NNge**

class Si IF : muy\_alta in {false,true} ^ alta in {false,true} ^ baja in {false} ^ muy\_baja in {false,true} ^ catedra\_cpf in {P} ^ catedra\_cpg in {P} ^ catedra\_pv in {P,N} ^ carrera in {Ing.Elect y Telec,Ing.Sis,Ing.Electm,Ing.Geo y OT,Ing.Geo y OT\_e,Tecnol.Elect y CI} (134)

class Si IF : muy\_alta in {true} ^ alta in {false,true} ^ baja in {false} ^ muy\_baja in {false,true} ^ catedra\_cpf in {N} ^ catedra\_cpg in {P,N} ^ catedra\_pv in {P} ^ carrera in {Ing.Elect y Telec,Ing.Sis,Tecnol.Elect y CI} (13)

class No IF : muy\_alta in {false,true} ^ alta in {false,true} ^ baja in {false,true} ^ muy\_baja in {false,true} ^ catedra\_cpf in {N} ^ catedra\_cpg in {N} ^ catedra\_pv in {N} ^ carrera in {Ing.Elect y Telec,Ing.Sis,Tecnol.Elect y CI} (52)

class No IF : muy\_alta in {false} ^ alta in {false,true} ^ baja in {false,true} ^ muy\_baja in {false,true} ^ catedra\_cpf in {N} ^ catedra\_cpg in {N} ^ catedra\_pv in {N} ^ carrera in {Ing.Electm,Ing.Geo y OT\_e} (10)

De las reglas generadas por el algoritmo se tomó las más relevantes existen 134 docentes que tienen una evolución de cátedra ya que en sus periodos de evaluación tienen una calificación positiva en capacidad profesional y capacidad pedagógica, cabe destacar que una segunda regla evidencia que la práctica de valores es muy importante, 13 docentes que obtuvieron una calificación final muy alta en algún periodo e incluso una calificación muy baja en otro periodo de evaluación, capacidad pedagógica o profesional negativa pero si ha obtenido una continua práctica de valores positiva entonces si tendrá una evolución de cátedra.

Cabe recalcar que para este punto se toman las calificaciones que un estudiante emite hacia un docente; la verificación de cursos, entrega de certificados, investigaciones que un docente ha realizado, entre otros, le compete a directivos de cada carrera. Por lo tanto la evolución de cátedra presentada en la presente investigación, es como la palabra mismo lo denota, una evolución en la enseñanza dentro del aula, la misma que puede servir como anexo a los procesos llevados a cabo por cada coordinación de carrera en la

definición y toma de decisiones en cuanto a un docente finalmente obtiene una evolución de cátedra.

### **Análisis Evolución de Cátedra**

Los resultados en este aspecto son los siguientes:

- ✓ En todas las carreras, un docente con evolución de cátedra por lo general tendrá una capacidad pedagógica y práctica de valores positiva obtenida de cada asignatura impartida.
- ✓ Las carreras con más docentes con evolución de cátedra son los que pertenecen a las carreras Ingeniería en Geología y Ordenamiento Territorial presencial (especial), la carrera de Tecnología en Electricidad y Control Industrial y Electromecánica con un 78.72%, 62.5%, 78.57%, 68.91% respectivamente.
- ✓ Un docente que pertenece a la carrera de Ingeniería en Electromecánica, en sus periodos de evaluación posee una calificación muy alta, pero una capacidad pedagógica o profesional negativa, y una práctica de valores positiva entonces no tiene una evolución de cátedra.
- ✓ Cuando un docente pertenece a la carrera de Ing. en Electrónica y Telecomunicaciones y en alguna evaluación posee una capacidad pedagógica o profesional negativa pero una constante práctica de valores positiva entonces si tiene una evolución de cátedra
- ✓ Si un docente ha obtenido una calificación Negativa dentro de un periodo y en un siguiente periodo se da una ausencia del mismo entonces no tendrá evolución de cátedra.
- ✓ La carrera que presenta más docentes sin evolución de cátedra es la carrera de Ingeniería en Sistemas y la carrera de Ingeniería en Electrónica y Telecomunicaciones con un 40.47%, y 45.94% respectivamente del total de docentes evaluados.
- ✓ Los docentes que pertenecen a las carreras de Ingeniería en Electrónica y Telecomunicaciones e Ingeniería en Electromecánica tienen una evolución de cátedra con mayor influencia en la capacidad pedagógica calificada dentro de la evaluación.

- ✓ Los docentes que pertenecen a la carrera de Ingeniería en Geología y Ordenamiento Territorial presencial o (especial), tienen una evolución de cátedra con mayor influencia en la capacidad profesional calificada dentro de la evaluación.
- ✓ En la carrera de ingeniería en Sistemas la mayor influencia en cuanto a la evolución de cátedra que tenga un docente dependerá de su calificación estrictamente en práctica de valores. Esto evidencia que los estudiantes ejercer una evaluación más crítica y consiente dentro de este componente de evaluación, es decir que un docente puede tener negativo en cuanto a su capacidad profesional, pero si se mantiene de forma constante con una práctica de valores positiva pues tendrá una evolución de cátedra.
- ✓ Un docente que no ha tenido una evolución de cátedra por lo general tendrá una calificación de Muy baja o Baja, en una o varias evaluaciones de desempeño y negativo en alguno de los componentes evaluados de forma constante.
- ✓ Docentes que obtuvieron una calificación final muy alta en algún periodo e incluso una calificación muy baja en otro periodo de evaluación, capacidad pedagógica o profesional negativa pero si ha obtenido una continua práctica de valores positiva entonces si tendrá una evolución de cátedra.
- ✓ Los docentes que pertenecen a la carrera de Tecnología en Electricidad y Control Industrial, al igual que la carrera de Ingeniería en Sistemas tienen una evolución de cátedra con mayor influencia en práctica de valores calificada dentro de la evaluación.

#### **4.2.1.1.3 Resultados en el Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.**

- **Modelo de Regresión Logística.**

En este aspecto se realizó un primer paso de entrenamiento para observar y generar un modelo que prediga la calificación que obtendrá un docente en base a las notas que tiene un estudiante, para luego aplicarlo a nuevos datos y predecir la instancia propuesta; cabe destacar que los datos son tomados aleatoriamente por el modelo y especificando una aplicación del modelo a un 30% de nuevos datos.

Es necesario señalar que para la aplicación del presente modelo se lo pudo también obtener con nuevos datos del proceso de evaluación, los mismos que no fueron otorgados debido a un cambio en la administración de la comisión interna de evaluación (Ver Anexo 9).

### **Predicción de la calificación de un docente en el proceso de evaluación de desempeño.**

En la siguiente figura (Ver Figura 81) se presenta la predicción de si un docente obtendrá una calificación Alta o Baja en base a las notas individuales de cada estudiante.

Row No.	confidence(Baja)	confidence(Alta)	prediction(calif_cualitativa)	nota
3290	0.121	0.879	Alta	8
3291	0.121	0.879	Alta	8
3292	0.121	0.879	Alta	8
3293	0.989	0.011	Baja	1.800
3294	0.019	0.981	Alta	9.900
3295	0.101	0.899	Alta	8.200
3296	0.282	0.718	Alta	7
3297	0.092	0.908	Alta	8.300
3298	0.076	0.924	Alta	8.500
3299	0.450	0.550	Alta	6.300
3300	0.121	0.879	Alta	8
3301	0.017	0.983	Alta	10
3302	0.776	0.224	Baja	4.920
3303	0.036	0.964	Alta	9.250
3304	0.076	0.924	Alta	8.500
3305	0.076	0.924	Alta	8.500
3306	0.843	0.157	Baja	4.500
3307	0.926	0.074	Baja	3.700
3308	0.984	0.016	Baja	2.200
3309	0.092	0.908	Alta	8.300
3310	0.092	0.908	Alta	8.300
3311	0.654	0.346	Baja	5.500
3312	0.700	0.300	Baja	5.300
3313	0.502	0.498	Baja	6.100
3314	0.083	0.917	Alta	8.400
3315	0.113	0.887	Alta	8.080
3316	0.177	0.823	Alta	7.580
3317	0.165	0.835	Alta	7.660
3318	0.122	0.878	Alta	7.990

Figura 81: Predicción de calificaciones en el proceso de evaluación de desempeño.

El modelo de la figura anterior muestra que si un estudiante por lo general tiene un promedio de nota superior a 7, mayor probabilidad tiene el docente de obtener una calificación de Alta. De forma inversa sucede lo mismo ya que por lo general los estudiantes que obtienen una nota menor a 6 tienden a calificar más bajo a sus docentes. Se debe señalar también que como en todo proceso existen excepciones a la regla que se puede tomar en cuenta.

Ya que la muestra tomada para la predicción fue del 30% de la totalidad de los datos es decir 3451 estudiantes por asignatura, en la (Figura 82) se presentan las instancias clasificadas según las categorías de predicción del modelo generado.

Role	Name	Type	Statistics	Range	Missings
confidence_Baja	confidence(Baja)	real	avg = 0.182 +/- 0.208	[0.017 ; 0.998]	0
confidence_Alta	confidence(Alta)	real	avg = 0.818 +/- 0.208	[0.002 ; 0.983]	0
prediction	prediction(calf_cualitativa)	nominal	mode = Alta (3180), l	Baja (271), Alta (3180)	0

Figura 82: Número de estudiantes que califican por categoría.

Se predice que 271 estudiantes califican con una calificación Baja a un docente, en base a sus notas y la mayor parte de estudiantes que corresponde a 3180 califica con una calificación Alta, igualmente en base a las notas obtenidas en cada asignatura.

- **Modelo ID3**

Con la generación de este modelo se pudo clasificar los datos mediante la aplicación del algoritmo seleccionado, de tal manera que el mismo permitió la descripción y observación de lo que ocurre actualmente en el proceso de evaluación docente (*patrones de comportamiento*).

En la siguiente figura (Ver Figura 83-85) se puede observar la clasificación de calificaciones docentes.

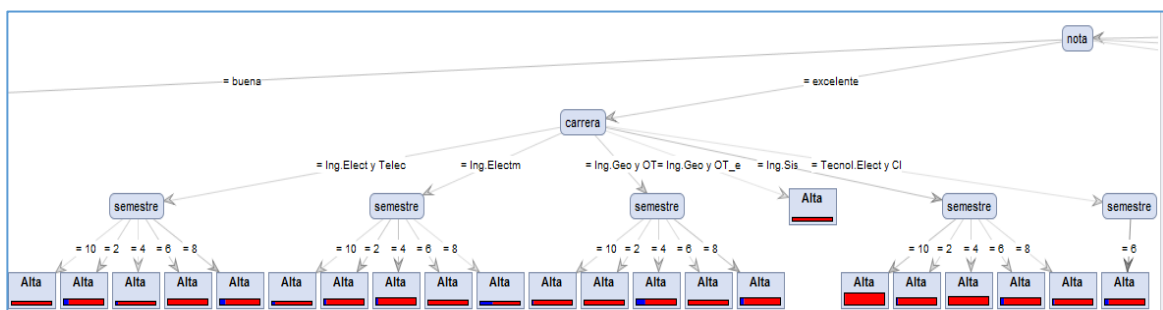


Figura 83: Clasificación calificación docente periodo académico 1.

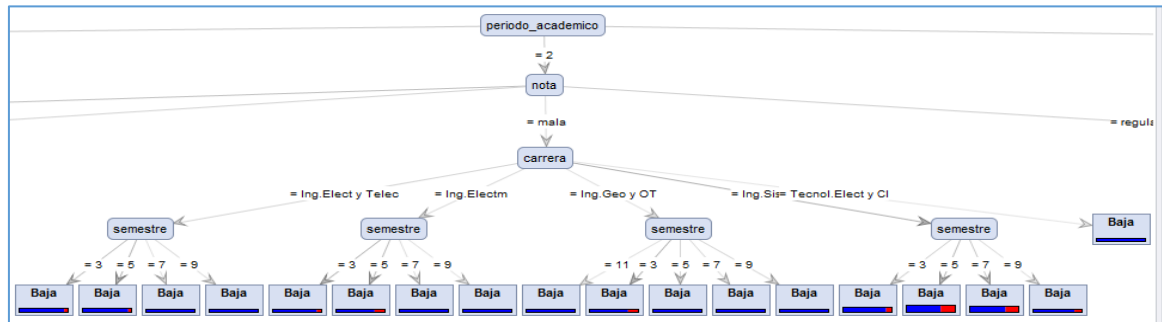


Figura 84: Clasificación calificación docente periodo académico 2.

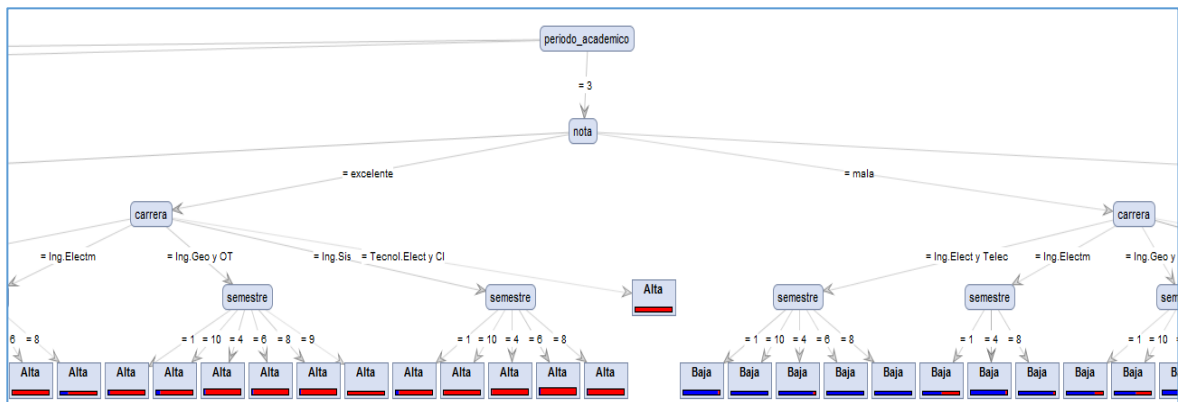


Figura 85: Clasificación calificación docente periodo académico 3.

En la figura se puede observar que el algoritmo clasifica las calificaciones que un docente obtiene en base a las notas que pertenece a los estudiantes. Es notorio observar un patrón de comportamiento en los 3 periodos académicos: es decir el árbol genera 3 ramificaciones par a los 3 periodos y en cada una de ellas se evidencia la influencia que tienen las notas de los estudiantes en la calificación de desempeño de un docente.

- **Modelo C4.5**

El comportamiento obtenido C4.5 a diferencia del modelo ID3, es que clasifica los datos tomando como base las carreras haciendo una comparación de las notas obtenidas por los estudiantes en cada asignatura y subdividiendo las mismas por periodos académicos, en este modelo no se toma los semestres debido a que el árbol lo toma como nodo raíz haciendo difícil la interpretación del mismo. (Ver Figura 86)

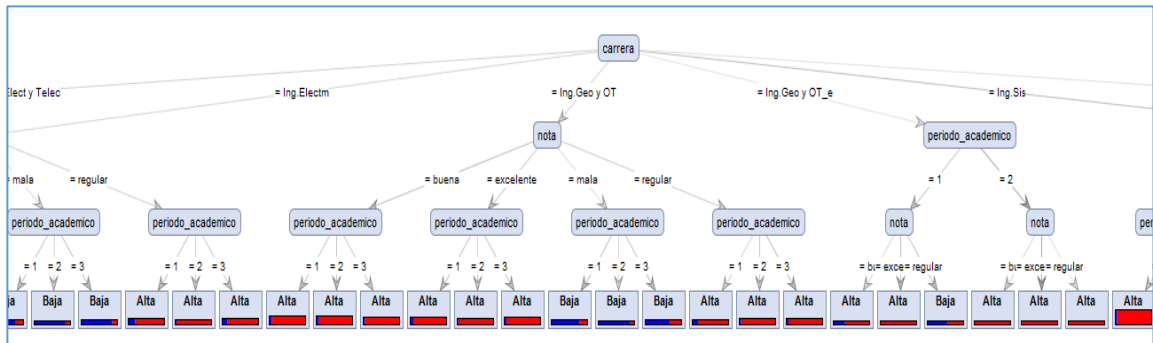


Figura 86: Clasificación modelo C4.5

**Patrones de Comportamiento:** el modelo refleja que a mayor nota de un estudiante en una asignatura, mayor probabilidad tiene un docente de obtener una buena calificación de desempeño y de igual manera a menor nota de un estudiante menor la probabilidad existe que un docente tenga una calificación de desempeño alta.

Es necesario resaltar que no todos los modelos pueden ser usados para predecir sino también para clasificar (Ver Figura 87) y obtener un mejor panorama de lo que sucede en una base de datos.

Por ende con el presente modelo se refuerza lo descrito anteriormente con el modelo de regresión e ID3 a mayor nota de un estudiante mejor calificación docente.

Role	Name	Type	Sum	Missings
id	cedula_docente	nominal	?	0
label	calf_cualitativa	nominal	?	0
confidence_Baja	confidence(Baja)	real	2135.521	19
confidence_Alta	confidence(Alta)	real	9351.479	19

Figura 87: Clasificación calificación docente por categorías.

En la figura anterior se puede observar que de la totalidad de registros 11506, 2135 estudiantes en todas las asignaturas califican Bajo, y la mayor cantidad de estudiantes en las asignaturas recibidas 9351 califican Alto a un docente dependiendo de la nota obtenida en cada asignatura. Dando un total de 19 registros sin clasificación debido al valor nulo en el atributo nota de la estructura analizada.



- **Análisis Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente**

Los resultados generales de los algoritmos aplicados en este aspecto son los siguientes:

- ✓ Si un estudiante sobrepasa la una calificación de 6, mayor probabilidad existe que un docente tenga una calificación alta.
- ✓ Si un estudiante tiene una calificación buena e incluso excelente, mayor será la probabilidad en ítems de calificación de 4 o 3 hacia un docente.
- ✓ En la mayoría de los casos si un estudiante tiene una nota menor a 6, menor será la probabilidad que un docente tenga una calificación alta.
- ✓ Por lo general un estudiante con una nota baja, califica con un porcentaje bajo y por ende un docente obtendrá como resultado una categoría de calificación en desempeño baja.

#### **4.2.1.2 Evaluación de los resultados de la minería de datos con respecto a Factores Críticos del Éxito.**

- **Primer Factor Crítico**

Como primer factor crítico de éxito están los criterios de interés que pueden existir dentro del proceso de evaluación, el poder observar los criterios de cada estudiante en cada asignatura por cada periodo académico, en cada área y carrera, puede ser muy tedioso por ende se establece un modelo que optimice aquel proceso e identifique en base a una estructura la existencia o no de criterios de interés.

Los algoritmos ID3 y CHAID arrojaron buenos resultados por ende:

- Los resultados obtenidos fueron del 92.29% para docentes que presentan un buen rendimiento en toda el área de la energía dentro del proceso, y 7.71% para la existencia de docentes con criterios de interés.

- **Segundo Factor Crítico**

El segundo factor crítico de éxito se centra en la evolución de cátedra que un docente puede tener a través de los periodos académicos, de ejercicio de sus funciones en su espacio de cátedra. Para el logro del mismo se elaboró al igual que los modelos anteriores una estructura que contenga el total de docentes de toda el AEIRNNR.

- Los modelos que mejor se desempeñaron fueron: Naive Bayes, ID3 y NNge.
- El modelo ID3 permitió describir el comportamiento de los datos en la existencia de la evolución de cátedra que pueda tener un docente a través de los años, esto basándose en los ítems más altos con que un estudiante califica a un docente en un periodo de evaluación, la clasificación corresponde al 67.18% de docentes con evolución de cátedra en los periodos de estudio, y un 32.81% de docentes sin dicha evolución.

- **Tercer Factor Crítico**

El tercer y último factor crítico de éxito se determinó la generación de un modelo que permita observar el comportamiento de las calificaciones docentes en base a las notas individuales de los estudiantes. Para esto se elaboró una estructura que posea la calificación docente y las notas de los estudiantes en todas las asignaturas de las carreras del AEIRNNR.

Los algoritmos con mejor rendimiento en este aspecto de análisis fueron Regresión logística, árbol de clasificación ID3 y C4.5.

- El modelo de regresión logística al igual que el modelo ID3 permitió la clasificación total de las instancias dando un 18.98% en la categoría de calificación Baja y un 81.01% de calificación alta.
- El modelo C4.5 permitió la clasificación en un 18.37% de calificación Baja y un 81.48% de calificación alta.

## **5. FASE 5: Culminación de la redacción de la memoria final del Proyecto Fin de Carrera.**

### **5.1 Redacción de un artículo científico basado en el estándar IEEE.**

Todo lo realizado en el presente trabajo de titulación se plasmó dentro de la memoria final, la misma que contiene los procesos y actividades realizados para la culminación exitosa de la investigación, pero además se elaboró un artículo científico basado en los estándares IEEE (Ver Anexo 10), para posteriormente enviarlo para su publicación dentro del Segundo Encuentro de Tecnologías de la Información y la Comunicación de las Universidades del Ecuador – TIC. EC 2014, llevado a cabo en la Universidad de Cuenca. Previo a esto el presente trabajo de titulación fue seleccionado dentro del II Concurso de Reconocimiento a la Investigación Universitaria Estudiantil: Galardones Nacionales 2014.

## **g. Discusión**

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

La propuesta alternativa describe el proceso realizado para el cumplimiento de los objetivos.

- **Analizar los datos almacenados, en el proceso de Evaluación de Desempeño Docente.**

Para la realización y alcance de este objetivo se llevó a cabo conversaciones con el responsable de la Unidad de Telecomunicaciones con respecto a la disposición de los datos para su posterior utilización. Además se realizó una exposición de la propuesta de investigación, a la coordinadora del proceso de evaluación de desempeño docente y a la responsable de Estadística y Base de Datos de la Unidad de Evaluación institucional, esto con el fin de evaluar la propuesta, de tal manera que no exista riesgo alguno en cuanto a la confidencialidad e integridad de los datos existentes en dicho proceso.

Luego de la obtención del consentimiento por parte de la coordinadora del proceso de evaluación, para poder acceder a los datos almacenados, se obtuvo el permiso respectivo para el acceso al web services de la Universidad con el propósito de obtener los datos en cuanto a las notas de los estudiantes en los periodos académicos de estudio (Ver Anexo 2, 3).

Ya con los permisos obtenidos se procedió al análisis de los datos disponibles con el fin de conocer la estructura, características y cantidad de los mismos.

Todo lo especificado anteriormente, se detalla en la Fase 1 de la sección Resultados.

- **Comparar y examinar las técnicas de minería de datos que mejor se ajustan al contexto del problema.**

En la realización del presente objetivo primero se recolectaron casos de éxito en donde se haya aplicado minería de datos dentro de la educación. Estos permitieron observar y obtener una idea más clara de las técnicas que se han utilizado y los resultados obtenidos en dichos casos de estudio. (Ver Capítulo I de Revisión, Literaria)

Luego se describió cada una de las técnicas y herramientas utilizadas en cada caso de éxito, para posteriormente comparar y examinar las diversas técnicas de minería de datos, de tal manera que se pueda visualizar desde una mejor perspectiva las técnicas más útiles dentro del problema abordado y como aportaron las mismas en la ejecución del trabajo investigativo (Ver Sección de Resultados: Fase 2: Selección de Técnicas).

- **Evaluar la efectividad de los procesos llevados a cabo, mediante las técnicas seleccionadas y del modelo propuesto.**

Para evaluar la efectividad de los procesos que se llevaron a cabo se siguió cada una de las fases descritas dentro de la metodología aplicada, de tal manera que se pudo obtener en primera instancia la aplicación de algoritmos específicos, generación de modelos, luego la evaluación de los mismos y posteriormente el análisis de resultados, para cada uno de estos aspectos se describe a continuación:

**Aplicación de algoritmos:** se aplicaron los siguientes algoritmos seleccionados Regresión Logística, ID3, C4.5, CHAID, Naive Bayes, W-REPTree, W-PART, W-NNge.

**Generación de Modelos:** en base a la aplicación de los algoritmos anteriormente mencionados se elaboró modelos enfocados a predecir y describir comportamientos dentro de la base de datos. Los modelos sirvieron para describir y predecir el comportamiento de los estudiantes y docentes del AEIRNNR, modelos como:

- Modelo para clasificar a docentes que tengan un excelente desempeño o posean criterios de interés dentro de un periodo de evaluación.
- Modelo para describir y conocer la evolución de cátedra que ha tenido un docente dentro de los periodos de estudio.
- Modelo para describir y predecir el comportamiento de los estudiantes al momento de evaluar a un docente en base a las notas individuales de cada asignatura.

**Evaluación:** ya en este punto se pudo evaluar tanto los algoritmos como los modelos generados para cada objetivo del negocio dentro de la investigación.

**Análisis de Resultados:** Se pudo tener resultados en base a los algoritmos que mejor se desempeñaron y los modelos con los cuales se obtuvieron mejores resultados:

- **Criterios de Interés o buen desempeño docente:** ID3, CHAID
- **Evaluación de cátedra:** Naive Bayes, ID3, NNge
- **Influencia de notas:** Regresión Logística, ID3, C4.5

Para la obtención y visualización de las opiniones de los estudiantes dentro del proceso se hizo uso de la herramienta Many Eyes.

Todo lo realizado dentro del presente objetivo se encuentra detallado en la Fase 3 Sección Resultados

- **Evaluar el modelo dentro de un contexto (escenario).**

Para el desarrollo del presente objetivo se tomó como escenario la Universidad Nacional de Loja (Ver Anexo 8), ya que la disposición y adquisición de los datos pertenecen a dicha entidad, el análisis fue abarcado en el AEIRNNR en las carreras de:

- Ingeniería en Electromecánica,
- Ingeniería en Geología Ambiental y Ordenamiento Territorial presencial y especial,
- Ingeniería en Electrónica y Telecomunicaciones y
- Tecnología en Electricidad y Control Industrial.
- Ingeniería en Sistemas

Los datos fueron proporcionados por la Unidad de Telecomunicaciones e Información (UTI), con la autorización de la Coordinación del Proceso de Evaluación de Desempeño Docente, datos correspondientes a los periodos académicos, 2011, 2012, 2013, puesto que hubo un cambio de administración en la comisión interna en la Evaluación de Desempeño Docente, los nuevos datos correspondientes al periodo 2014 no fueron otorgados para su respectivo análisis (Ver anexo 9)

La ejecución del presente trabajo investigativo dentro de este escenario permitió conocer y describir la situación del proceso de evaluación de desempeño docente, pero sobre todo y en base al análisis realizado, presentar un modelo útil en a la toma de decisiones académicas como administrativas (Ver Anexo 10).

Todo lo mencionado en este apartado se lo puede evidenciar en la Sección Resultados: Fase 4 Evaluación del modelo generado e interpretación de Resultados

Con lo argumentado anteriormente se puede concluir que el presente trabajo de titulación tuvo una culminación exitosa, debido al cumplimiento y alcance de cada uno de los objetivos planteados.

Además se puede afirmar que los resultados obtenidos a través del presente trabajo de titulación pueden generar nuevas líneas de investigación que refuercen o planteen nuevas técnicas o métodos de análisis en instituciones educativas.

## **2. Valoración Técnica Económica Ambiental**

El presente trabajo de titulación denominado “Análisis del proceso de Evaluación al Desempeño Docente, aplicando Técnicas de Minería de Datos”, da como resultado la descripción de comportamientos y obtención de patrones de interés dentro del proceso de evaluación, como la influencia que tienen las notas de los estudiantes al momento de evaluar a un docente, los criterios de interés o buen rendimiento que puede tener un docente dentro de un periodo de evaluación, conocer que ocurre o cuales son las opiniones de los estudiantes hacia un docente, el grado de ocurrencias de dichas opiniones en cada carrera, y la evolución de cátedra que ha tenido un docente dentro de los periodos académicos seleccionados.

Análisis que se considera un aporte a la educación superior pero sobre todo al mejoramiento en la calidad de la misma, debido a que los modelos generados pueden ser un apoyo a la toma de decisiones y su implantación ayudaría a la optimización de procesos administrativos, dicho en otras palabras el presente trabajo investigativo constituye una contribución técnica.

La valoración económica de la presente investigación radica, en el libre acceso de las herramientas utilizadas y la información disponible con la que se llevó la presente investigación la tecnología facilita herramientas de uso libre para el análisis de información con Minería de Datos, en donde se realizó la extracción de conocimiento basado en los datos existentes en las bases de datos de la universidad.

La justificación ambiental tiene su base en que no existió peligro alguno para el medio ambiente, al momento de realizar el análisis al desempeño docente mediante técnicas de minería de datos.

A continuación se detalla el talento humano, los bienes y servicios utilizados en el trabajo de titulación:

El talento humano que fue parte del Trabajo de Titulación, estuvo conformado por el investigador el cual ejecutó cada fase y objetivo propuesto dentro de la investigación, y el asesor o director, quien supervisó cada aspecto del desarrollo y aportó con sus conocimientos, de tal forma que se pueda llevar a cabo una investigación alcanzable y ordenada. En la siguiente tabla (Ver Tabla CV) se describe un aproximado en tiempo y costo para el término exitoso del Trabajo de Titulación.

**TABLA CV: TALENTO HUMANO**

<b>Equipo Trabajo</b>	<b>Tiempo (Horas)</b>	<b>Precio/ Hora (\$)</b>	<b>Valor Total (\$)</b>
Investigador	1400	5.00	7000.00
Asesor y Director	280	5.00	1400.00
SUBTOTAL (\$)			8400.00

Las tablas (Ver tabla CVI, CVII, CVIII, CIX), detallan los Recursos Hardware, Software, materiales y servicios que fueron utilizados a lo largo de todo el trabajo investigativo, cada recurso utilizado fue indispensable ya sea en la generación o diseño de los modelos, como en la ejecución de cada proceso de minería de datos. La tabla CVII describe el costo total de los recursos empleados y da un estimado del costo del proyecto elaborado.

**TABLA CVI: RECURSOS HARDWARE.**

<b>Hardware</b>	<b>Precio Unitario (\$)</b>	<b>Precio Total (\$)</b>
<b>1 Computador Personal.</b>	\$ 678,00	\$ 678,00
<b>1 Impresora</b>	\$ 56,00	\$ 56,00
<b>2 flash memory</b>	\$ 10.00	\$ 20,00
		754,00



**TABLA CVII: RECURSOS SOFTWARE.**

Software	Precio Unitario (\$)	Coste Total (\$)
<b>Rapid Miner</b>	\$ 0.00	\$ 0.00
<b>Pg Admin III</b>	\$ 0.00	\$ 0.00
		\$ 0.00

**TABLA CVIII: MATERIALES Y SERVICIOS**

<b>Materiales y Servicios</b>	<b>Descripción</b>	<b>Costo por Mes</b>	<b>Costo Desarrollo del Proyecto</b>
Materiales de oficina	Papel, esferos, carpetas, cartuchos, etc	20,00	200,00
Servicios	Luz	15,00	150,00
	Acceso a internet	20,00	200,00
	teléfono	10,00	100,00
	Transporte	7.00	70.00
<b>Total</b>			<b>720,00</b>

**TABLA CIX: COSTE TOTAL DEL PROYECTO**

Recursos	Coste Total (\$)
<b>Costes de Personal</b>	\$ 8400
<b>Costes de Hardware</b>	\$ 754,00
<b>Costes de Software</b>	\$ 0,00
<b>Costes de Material y Servicios.</b>	\$ 720,00
<b>Total:</b>	<b>\$ 9874,00</b>

La tabla CV detalla las horas invertidas en el desarrollo del trabajo de titulación las mismas que sobrepasan a las propuestas inicialmente en el anteproyecto. Esto debido a que en un principio no se obtuvo la disponibilidad de los datos, los mismos que se entregaron luego de tres meses de la aprobación del presente proyecto investigativo, acarreado consigo que se extienda la ejecución del mismo dos meses más de lo previsto, y trabajando a razón de 7 horas diarias de lo establecido inicialmente.

## **h. Conclusiones**

Las conclusiones obtenidas al finalizar el presente trabajo de Titulación son las siguientes:

- El análisis de diversas técnicas de minería de datos da una mejor perspectiva y más opciones de solución a un problema abordado. Se debe tomar en cuenta que pueden existir diversas técnicas para un problema y que cada algoritmo empleado en la generación de un modelo de minería requiere un formato concreto de datos.
- Los procesos que se llevan a cabo dentro de la minería de datos, muchas de las veces no es como lo presentan en blogs, foros o documentos no tan específicos, que denotan a la minería como un proceso sencillo que no demanda mucho conocimiento o entrenamiento. Al contrario la minería de datos exige un previo conocimiento de las técnicas útiles para un tipo de datos, demanda un entrenamiento constante de los modelos generados, requiere de la elaboración de estructuras para la posterior aplicación de algoritmos, un manejo y conocimiento de programación de bases de datos y técnicas estadísticas.
- La metodología CRISP-DM fue elegida para la realización del presente trabajo de titulación debido a que presenta en detalle los pasos para el desarrollo de cualquier proyecto de minería de datos, documentación disponible, pero sobre todo porque la misma es muy utilizada en el campo empresarial y académico.
- Los algoritmos para la generación de modelos y resultados se seleccionaron en base al rendimiento, validación y menor error presentado por cada uno, los cuales fueron: ID3, CHAID, Naive Bayes, NNge, Regresión Logística y C4.5.

## **i. Recomendaciones**

Una vez desarrollado y culminado el Trabajo de Titulación, es necesario presentar ciertas recomendaciones:

- En base a lo experimentado mediante el desarrollo del trabajo de titulación, se recomienda que previo a la validación de un modelo generado, se aplique varios algoritmos para de esta forma obtener una mejor perspectiva del comportamiento del modelo y por ende una mejor obtención de resultados.
- Se recomienda además invertir tiempo de calidad en la generación de las estructuras para no tener que regresar constantemente a la fase de preparación de los datos, ya que posteriormente se aplicarán los diversos algoritmos de minería de datos, y de las mismas dependerá el entrenamiento, diseño, rendimiento, comportamiento del modelo y generación de resultados.
- Se recomienda además ampliar el campo de estudio empleando minería de texto (texmining) en las opiniones dentro del proceso, ya que esto involucra nuevas fases de desarrollo, diferentes a la minería de datos, dando así un diferente tratamiento de los datos en cuanto a modelado o resultados, empleando técnicas y algoritmos en las opiniones existentes.
- Actualmente los datos en cuanto a la evaluación docente requieren un mejor tratamiento, ya que los mismos generan información valiosa constituyendo el activo más importante de la institución esto con el fin de elaborar estrategias que brinden incentivos cuando un docente presenten un excelente desempeño o buen desempeño de enseñanza, o ataquen las deficiencias de cátedra en docentes que lo requieran, por lo tanto se recomienda tomar en cuenta el modelo generado o a su vez extenderlo a las demás áreas dentro de la universidad.

## **j. Bibliografía**

### **Referencias Bibliográficas**

[1] Proyecto de Evaluación del Desempeño Docente. Universidad Nacional de Loja [En Línea]: <<http://xurl.es/nme19>>. [Acceso: 28-October-2013]

[2] Ley Orgánica de Educación Superior. [En Línea] <[http://www.utelvt.edu.ec/LOES\\_2010.pdf](http://www.utelvt.edu.ec/LOES_2010.pdf)>. [Acceso: 28-October-2013]

[3] La Evaluación de la Actividad Docente Del Profesorado en la Universidad Pública De Navarra. Universidad Pública de Navarra. Diciembre 2012 [En Línea] <[http://www.unavarra.es/digitalAssets/186/186243\\_metodoevaluacion.pdf](http://www.unavarra.es/digitalAssets/186/186243_metodoevaluacion.pdf)>. [Acceso: 28-October-2013]

[4] RIZO Héctor, Evaluación Del Docente Universitario, Una Visión Institucional. Universidad Autónoma de Occidente. [En Línea]: <<http://www.rieoei.org/deloslectores/883Rizo.pdf>>. [Acceso: 28-October-2013]

[5] MUÑOZ Jesús M., RÍOS María P., PAZ Eduardo, Evaluación Docente vs. Evaluación de la Calidad. (2002). Revista Electrónica de Investigación y Evaluación Educativa, v.8, n.2. [En Línea] <[http://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2\\_4.htm](http://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2_4.htm)>. [Acceso: 29-October-2013]

[6] TOLOSA Pamela, Evaluación de la calidad docente. Universidad Nacional del Sur. Revista sobre enseñanza del Derecho. n 7, 2006, ISSN 1667-4154, págs. 411-418.

[7] VALDÉS Héctor, Encuentro Iberoamericano sobre Evaluación del Desempeño Docente. Mayo 2000. [En Línea] <<http://www.oei.es/de/rifad01.htm>>. [Acceso: 29-October-2013]

[8] MOLINA, Luis C. La minería de datos: torturando a los Datos Hasta Que confiesen. [En Línea]. <<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>>. [Acceso: 29-October-2013]

[9] FAYYAD Usama, PIATETSKY Shapiro, Smyth Padhraic From Data Mining to Knowledge Discovery in Databases. 1996. [En Línea]. <<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>>. [Acceso: 29-October-2013]

[10] VALERO O. Sergio, Aplicación de Técnicas de Minería de Datos para Predecir Deserción. Universidad Tecnológica de Izúcar de Matamoros. 2009. [En Línea].

<<http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf>>. [Acceso: 29-October-2013]

[11] LUAN Jing, Aplicaciones de minería de datos en la educación superior. Mayo 2010. [En Línea]. <<http://public.dhe.ibm.com/common/ssi/ecm/es/imw14303eses/IMW14303ESES.PDF>>. [Acceso: 29-October-2013]

[12] JIMÉNEZ Álvaro, ALVAREZ Hugo, Minería de Datos en la Educación. Universidad Carlos III de Madrid. [En Línea]. <<http://www.it.uc3m.es/jvillena/irc/practicas/10-11/08mem.pdf>>. [Acceso: 1-Noviembre-2013]

[13] MÁRQUEZ Carlos, ROMERO Cristóbal, VENTURA Sebastián, Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos. [En Línea]. <<http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf>>. [Acceso: 1-Noviembre-2013]

[14] GONZÁLEZ Ernesto, PÉREZ Zady, ESPINOSA Ivet, ALVAREZ Susel, Obtención de Patrones y Reglas en el Proceso Académico de La Universidad de Las Ciencias Informáticas Utilizando Técnicas de Minería de Datos. Universidad de las Ciencias Informáticas. [En Línea] <<http://eprints.rclis.org/10937/1/Suzel.pdf>>. [Acceso: 1-Noviembre-2013]

[15] PULLA Cinthia, Técnicas de minería de datos para identificar patrones de colaboración de los estudiantes que hacen uso del EVA de la UTPL. Universidad Técnica Particular de Loja. [En Línea] <<http://cepra.utpl.edu.ec/handle/123456789/1013>>. [Acceso: 1-Noviembre-2013]

[16] HIDALGO Ruth, Análisis de las interacciones de los estudiantes de la UTPL con las herramientas del Eva, mediante el análisis de redes sociales y la aplicación de técnicas de minería de datos, para determinar indicadores de interacción basados en la colaboración. Universidad Técnica Particular de Loja. [En Línea] <<http://cepra.utpl.edu.ec/bitstream/123456789/1472/3/TesisRuth.pdf>>. [Acceso: 1-Noviembre-2013]

[17] ROMERO Cristóbal, VENTURA Sebastián, GARCÍA Enrique. Data mining in course management systems: Moodle case study and tutorial. Universidad de Córdoba. [En Línea] <<http://goo.gl/gN52t>>. [Acceso: 5-Noviembre-2013]

[18] GONZÁLEZ José, Generalidades de la Minería de Datos, Universidad de la Laguna [EnLínea].<[http://www.ecofin.ull.es/users/jggomez/D%20Bdr\\_Erp/6%20Mineria/Mineria.pdf](http://www.ecofin.ull.es/users/jggomez/D%20Bdr_Erp/6%20Mineria/Mineria.pdf)>. [Acceso: 5-Noviembre-2013]

- [19] VALERO O. Sergio, Aplicación de Técnicas de Minería de Datos para Predecir Deserción. Universidad Tecnológica de Izúcar de Matamoros. 2009. [En Línea]. <<http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf>>. [Acceso: 17-Noviembre-2013]
- [20] ROMÁN G. Ayesha, Minería de Datos en Encuestas de Profesores al Fin de Semestre de La Facultad de Ingeniería, Unam. Universidad Nacional Autónoma de México. Enero 2011. [En Línea]. <<http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/227/tesis.pdf?sequence=1>>. [Acceso: 17-Noviembre-2013]
- [21] GONZÁLEZ Juan, Sistema de Apoyo para la Acreditación de la Calidad de Programas Académicos de la Universidad de Caldas, Aplicando Técnicas en Minería De Datos. Universidad Autónoma de Manizales. Noviembre 2011. [En Línea] <[http://repositorio.autonoma.edu.co/jspui/bitstream/11182/350/1/Msc.GyDlloSoft\\_InformeFinal\\_JuanCarlosGonzalez.pdf](http://repositorio.autonoma.edu.co/jspui/bitstream/11182/350/1/Msc.GyDlloSoft_InformeFinal_JuanCarlosGonzalez.pdf)>. [Acceso: 17-Noviembre-2013]
- [22] Análisis de Decisiones II: Árboles de decisión. Universidad Tec. Milenio. [En Línea]. <<http://www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r94433.PDF>>. [Acceso: 17-Noviembre-2013]
- [23] Fernando Berzal Galiano, Un método alternativo para la construcción de árboles de decisión. Universidad de Granada. Junio 2002. [En Línea]. <<http://goo.gl/8c9WRf>>. [Acceso: 17-Noviembre-2013]
- [24] C. Blanco, F. García, Minería de Datos: Predicción de las condiciones meteorológicas [En Línea]. <<http://www.it.uc3m.es/~jvillena/irc/practicas/05-06/14mem>>. [Acceso: 17-Noviembre-2013]
- [25] R. Villarreal, Modelo de referencia CHISP-DM para el desarrollo de proyectos de minería de datos y CHAID como técnica de análisis para obtener información en Minería de Datos. InterSedes, 2012, vol. 13, no 25.
- [26] C. Tirnauca, Minería de Datos: Arboles de decisión. Universidad de Cantabria. 2012. [EnLínea].<<http://personales.unican.es/tirnaucac/Slides/121120%20Teoria%20%28handout%29.pdf>>. [Acceso: 17-Noviembre-2013]
- [27] L. Sucar, Redes Bayesianas. [En Línea]. <<http://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>>. [Acceso: 17-Noviembre-2013]
- [28] P. Larrañaga, Redes Bayesianas. Universidad del País Vasco. [En Línea].<<http://www.sc.ehu.es/ccwbayes/publications/postscript/lalaguna260702.pdf>>. [Acceso: 17-Noviembre-2013]

- [29] A. F. Bonilla, M. A. Ojeda, Fabricio Echeverría, Implementación de Minería de Datos Basada en Redes Bayesianas para La toma de Decisiones en Los Registros Académicos. [En Línea]< <<http://www.dspace.espol.edu.ec/bitstream/123456789/679/1/1193.pdf>>. [Acceso: 17-Noviembre-2013]
- [30] F. Izaurieta, C. Saavedra, Redes Neuronales Artificiales. Universidad de Concepción Chile. [En Línea]. <<http://www.uta.cl/charlas/volumen16/Indice/Ch-csaavedra.pdf>>. [Acceso: 17-Noviembre-2013]
- [31] J. Rojo, Regresión con variable dependiente cualitativa. Instituto de Economía y Geografía. 2007. [En Línea].< [http://portal.uned.es/pls/portal/docs/PAGE/UNED\\_MAIN/LAUNIVERSIDAD/VICERRECTO RADOS/INVESTIGACION/O.T.R.I/OFERTAS%20TECNOLOGICAS/DMAC/DOCUMENTO S%20Y%20TUTORIALES/REGRESION\\_VARIABLE\\_DEPENDIENTE\\_DICOTOMICA\\_3.P DF](http://portal.uned.es/pls/portal/docs/PAGE/UNED_MAIN/LAUNIVERSIDAD/VICERRECTO RADOS/INVESTIGACION/O.T.R.I/OFERTAS%20TECNOLOGICAS/DMAC/DOCUMENTO S%20Y%20TUTORIALES/REGRESION_VARIABLE_DEPENDIENTE_DICOTOMICA_3.P DF)>. [Acceso: 28-Noviembre-2013]
- [32] C. Márquez, C. Romero, S. Ventura, Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos. Noviembre 2012 [En Línea]. < <http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf>>. [Acceso: 28-Noviembre-2013]
- [33] G. Alberca, M. Agila, F. Loja, P. Valdiviezo, J. C. Jiménez, Recolección de datos de interacción de alumnos en una plataforma E-learning para obtener indicadores de interés de su actividad aplicando técnicas de aprendizaje automático. Universidad Técnica Particular de Loja [En Línea]< <http://www.iiis.org/CDs2009/CD2009CSC/SIECI2009/PapersPdf/X036SN.pdf> > [Acceso: 28-Noviembre-2013]
- [34] A. Rodríguez, J. Martínez, J. Carrasco, J. Ruiz, Minería de Reglas de Asociación sobre Datos Mezclados. Instituto Nacional de Astrofísica, Óptica y Electrónica, Centro de Aplicaciones de Tecnologías de Avanzada. [En Línea]< <http://ccc.inaoep.mx/portalfiles/file/CCC-09-001.pdf>>. [Acceso: 28-Noviembre-2013]
- [35] G. Bressán, Almacenes de Datos y Minería De Datos. Julio 2003. [En Línea]<<http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/MineriaDatosBressan.htm#Técnicas de MD>>. [Acceso: 28-Noviembre-2013]
- [36] R. CRESPO, J. VILLENA, Minería de Datos. [En Línea]. <[http://www.it.uc3m.es/rcrespo/docencia/irc/apuntes/Mineria\\_de\\_datos.pdf](http://www.it.uc3m.es/rcrespo/docencia/irc/apuntes/Mineria_de_datos.pdf)>. [Acceso: 28-Noviembre-2013]

- [37] J. Cubero, F. Berza, Herramientas de Minería de Datos, Universidad de Granados. [En Línea]<<http://elvex.ugr.es/decsai/intelligent/workbook/D1%20KNIME.pdf>>. [Acceso: 30-Noviembre-2013]
- [38]KNime.[EnLínea]<<http://www.exa.unicen.edu.ar/catedras/dmining/clases/PresentacionKNime.pdf>>. [Acceso: 30-Noviembre-2013]
- [39] C. Corso, F. Gibellini, Uso de herramienta libre para la generación de reglas de asociación, facilitando la gestión eficiente de inventarios e incidentes. Universidad Tecnológica Nacional. F.R.C. [En Línea] <<https://sl.linti.unlp.edu.ar/wp-content/uploads/2012/08/GeneracionReglasAsociacion.pdf>>. [Acceso: 30-Noviembre-2013]
- [40] L. Guerra. Primeros pasos con Knime. Mayo 2008. [En Línea]. <[https://laurel.datsi.fi.upm.es/\\_media/docencia/cursos/inap/ejemplodm.pdf](https://laurel.datsi.fi.upm.es/_media/docencia/cursos/inap/ejemplodm.pdf)>. [Acceso: 30-Noviembre-2013]
- [41] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. Witten, and L. Trigg, “Weka-a machine learning workbench for data mining,” in Data Mining and Knowledge Discovery Handbook, Springer, 2010, pp. 1269–1277.
- [42] D. Garcia M., Manual de Weka. [En Línea]. <<http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>>. [Acceso: 30-Noviembre-2013]
- [43] S. R. Garner, “Weka: The waikato environment for knowledge analysis,” in Proceedings of the New Zealand computer science research students conference, 1995, pp.57–64.
- [44] CUBERO Juan C., BERZA Fernando, HERRRERA F. Fundamentos de Minería de Datos. Universidad de Granada. [En Línea]. <<http://sci2s.ugr.es/docencia/m1/Preprocesamiento-Weka-MD.pdf>>.
- [45] F. García, Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de MedioAmbiente (CEAMA). Universidad de Granada. 20113. [En Línea]. <<http://goo.gl/AFJDCC>>. [Acceso: 30-Noviembre-2013]
- [46] A. Mori, Medir el beneficio económico y en los procesos con la aplicación de un sistema de minería de Datos para la empresa DIPESA (Descripción de Herramientas). Universidad Tecnológica del Perú. 2012. [En Línea]<<http://goo.gl/JKiRXN>>. [Acceso: 01-Diciembre-2013]



- [47] RapidMiner User Guide. Marzo 2009. [En Línea]<[http://www.is.informatik.uni-duisburg.de/courses/im\\_ss09/uebung/rapidminer-4.4-tutorial.pdf](http://www.is.informatik.uni-duisburg.de/courses/im_ss09/uebung/rapidminer-4.4-tutorial.pdf)>. [Acceso: 01-Diciembre-2013]
- [48] MIERSWA I., WURST M., R., KLINKENBERG M., YALE: Rapid Prototyping for Complex Data Mining Tasks, in Proc. 12th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining (KDD-06) (2006) 935–940
- [49] R. Wirth, J. Hipp, CRISP-DM: Towards a Standard Process Model for Data Mining [En línea]:<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf>>. [Acceso: 01-Diciembre-2013]
- [50] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth CRISP-DM 1.0: Step-by-step data mining guide. [En línea]:<<ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>>. [Acceso: 01-Diciembre-2013]
- [51] IBM, Manual CRISP-DM de IBM SPSS Modeler. [En línea]:<<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>>. [Acceso: 01-Diciembre-2013]
- [52] J. Gallardo, Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM. 2009. [En Línea] <<http://goo.gl/4JCF1a>>. [Acceso: 01-Diciembre-2013]
- [53] E. Fischer, Modelo Para La Automatización Del Proceso De Determinación De Riesgo De Deserción En Estudiantes Universitarios. Universidad De Chile. [En Línea]<<http://preu.unillanos.edu.co/sites/default/files/fields/documentos/PREDICION%20DESERCI.pdf>>. [Acceso: 01-Diciembre-2013]
- [54] C. Pérez, La reforma educativa a la luz del cambio de paradigma productivo. 2000. [En Línea]. <[http://www.carlotaperez.org/downloads/pubs/Refeduc\\_OREALC.pdf](http://www.carlotaperez.org/downloads/pubs/Refeduc_OREALC.pdf)>.
- [55] J. Salmi, Educación Superior Enfrentando los Retos del Siglo XXI [En Línea]. <<http://goo.gl/GeQ6Cp>>
- [56] R. Ramírez, MINTEGUIAGA Analía, “Transformaciones En La Educación Superior Ecuatoriana: Antecedentes Y Perspectivas Futuras Como Consecuencia De La Nueva Constitución Política”. Revista Ess, vol. 15, no. 1, pp 131-148, 2010.
- [57] R. Ramírez, Tercera Ola De Transformación De La Educación Superior En Ecuador. 2013. [En Línea]. <<http://www.educacionsuperior.gob.ec/wp-content/uploads/downloads/2013/08/Tercera-ola-de-transformaci%C3%B3n-de-la-educaci%C3%B3n-superior-en-Ecuador3.pdf>>. [Acceso: 15-Diciembre-2013]

- [58] Constitución del Ecuador. [En Línea]. <[http://www.asambleanacional.gov.ec/documentos/constitucion\\_de\\_bolsillo.pdf](http://www.asambleanacional.gov.ec/documentos/constitucion_de_bolsillo.pdf)>. [Acceso: 15-Diciembre-2013]
- [59] Reglamento De Evaluación Del Desempeño Del Personal Académico De La Universidad De Cuenca. [En Línea] Disponible en: <<http://www.ucuenca.edu.ec/images/stories/file/Ucuenca/Reglamentos/REGLAMENTO%20EVALUACION%20DOCENTE.pdf>>. [Acceso: 15-Diciembre-2013]
- [60] S. Martínez, L. Larios, Revisión del proceso de evaluación docente por los estudiantes, Universidad de Colima. Febrero de 2012. [En Línea] <<http://www.ucol.mx/personalacademico/documentos/evaluacionDocente.pdf>>. [Acceso: 15-Diciembre-2013]
- [61] A. Morán, Evaluación de la calidad Educativa del desempeño docente y Directivo en el Colegio Fiscal mixto “Dr. Luis Espinoza Tamayo” del recinto Villanueva del cantón Naranjal provincia del Guayas en la Educación Básica y Bachillerato del Ecuador durante el año académico 2012-2013. Universidad Técnica Particular de Loja. 2013. [En Línea] <<http://dspace.utpl.edu.ec/bitstream/123456789/7151/1/Moran%20Ramirez%20Arminda%20Maritza.pdf>>. [Acceso: 15-Diciembre-2013]
- [62] E. Monterroso, Normas ISO: ISO 9000. Universidad Nacional de Luján. [En Línea] <<http://www.unlu.edu.ar/~ope20156/normasiso.htm>>. [Acceso: 15-Diciembre-2013]
- [63] L. Moratal, Modelo De Abraham Flexner. Mayo 2011. [En Línea] <<http://investigacionydocencia-lmi.blogspot.com/2011/05/modelo-de-abraham-flexner.html>>. [Acceso: 15-Diciembre-2013]
- [64] R. Pesantes, Gerencia académica y diagnóstico de la gestión docente del Instituto superior Tecnológico ESCA. Propuesta de instrumentos de evaluación docente por competencias. Universidad De Guayaquil. [En Línea] <<http://repositorio.ug.edu.ec/bitstream/redug/1487/2/Tesis%20RAUL%20PESANTES%20part%202.pdf>>. [Acceso: 15-Diciembre-2013]
- [65] STUFFLEBEAM, D., SCHINKFIELD, A. El Método Evaluativo De Scriven Orientado Hacia El Consumidor. Universidad Nacional Abierta. [En Línea] <<http://postgrado.una.edu.ve/evaluacionaprendizajes/paginas/stufflebeamcap10.pdf>>. [Acceso: 15-Diciembre-2013]
- [66] Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC), Informe Sobre La Educación Superior En América Latina Y El Caribe. 2000-2005. Mayo 2006 [En Línea]. <<http://goo.gl/tPCSKI>>. [Acceso: 15-Diciembre-2013]

- [67] P. Selín, Repitencia y Deserción Universitaria en América Latina. Mayo 2012. [En Línea]. <<http://www.universidadfutura.org/documentos/repitencia-y-desercion-universitaria-en-america-latina/>>. [Acceso: 15-Diciembre-2013]
- [68] J. Calderón, Estudio Sobre La Repitencia Y Deserción En La Educación Superior De Guatemala. Septiembre de 2005. [En Línea]. <[http://www.alfaguia.org/alfaguia/files/1320941990\\_9453.pdf](http://www.alfaguia.org/alfaguia/files/1320941990_9453.pdf)>. [Acceso: 15-Diciembre-2013]
- [69] E. Hanushek. "The Trade-Off Between Child Quantity and Quality.", Journal of Political Economy Vol. 100, No. 1, pp. 84-117. 1992.
- [70] J. Rivers, W. Sanders. "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement". Research Progress Report, University of Tennessee Value-Added Research and Assessment Center, Knoxville, TN. 1996. [En Línea]. Disponible: <[http://www.cgp.upenn.edu/pdf/Sanders\\_Rivers-TVASS\\_teacher%20effects.pdf](http://www.cgp.upenn.edu/pdf/Sanders_Rivers-TVASS_teacher%20effects.pdf)>. [Acceso: 15-Diciembre-2013]
- [71] S. Vallejos, Que es la Minería de Datos, Universidad Nacional del Nordeste. [En línea]: <[http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria\\_Datos\\_Vallejos.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf)>. [Acceso: 20-Diciembre-2013]
- [72] J. Hernández, Minería de Datos: Otros Aspectos, Universidad Politécnica de Valencia. [En línea]: <<http://users.dsic.upv.es/~jorallo/master/dm5.pdf>>. [Acceso: 20-Diciembre-2013]
- [73] K. Sánchez, R. Ruiz, J Riquelme. Minería de datos: Conceptos y tendencias. Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial, 2006, vol. 10, no 29, p. 11-18.
- [74] A. Gómez, Inteligencia de negocios, una ventaja competitiva para las organizaciones, Revista CIENCIA Y TECNOLOGÍA, vol. 8, no. 22, pp. 85–96, 2013.
- [75] P. López, Qué es Inteligencia de Negocios (Business Intelligence). [En línea]: <<http://www.itmadrid.com/blog/que-es-inteligencia-de-negocios-business-intelligence/>>. [Acceso: 20-Diciembre-2013]
- [76] F. Mata, Árboles de decisión: Minería de Datos [En línea]: <<http://goo.gl/x30bR2>>. [Acceso: 20-Diciembre-2013]
- [77] J. Botía, Arboles de decisión en aprendizaje automático y minería de datos: Tratamiento Inteligente de la Información y Aplicaciones. Universidad de Murcia. Octubre 2007. [En línea]: <<http://goo.gl/MCa3xc>>. [Acceso: 12-Enero-2014]

- [78] M. Moreno, Luis Quintales, Francisco García, J. Polo, Aplicación de Técnicas de Minería de Datos en la Construcción y Validación de Modelos Predictivos y Asociativos a Partir de Especificaciones de Requisitos de Software, Universidad de Salamanca. [En línea]: <<http://www.sc.ehu.es/jiwdocoj/remis/docs/minerw.pdf>>. [Acceso: 12-Enero-2014]
- [79] Gutiérrez, J.M., Cofiño, A.S., Sordo, C.M., San-Martín, D., Bedia, J. Minería de Datos. Redes Bayesianas y Neuronales. [En línea]: <[http://www.meteo.unican.es/es/research/mineria\\_datos](http://www.meteo.unican.es/es/research/mineria_datos)>. [Acceso: 12-Enero-2014]
- [80] P. Felgae, Redes Bayesianas Aplicadas a Minería de Datos Inteligente. Facultad de Ingeniería. Universidad de Buenos Aires. [En línea]: < <http://laboratorios.fi.uba.ar/lsi/p-felgaeer-proyectedetesis.htm>>. [Acceso: 22-Febrero-2014]
- [81] E. Morales, Redes Bayesianas en Minería de Datos. [En línea]: < <http://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/node68.html>>. [Acceso: 22-Febrero-2014]
- [82] Redes neuronales artificiales para aprendizaje automático y minería de datos: Tratamiento Inteligente de la Información y Aplicaciones. Universidad de Murcia. Octubre 2007. [En línea]: <<http://goo.gl/YTAmML>>. [Acceso: 22-Febrero-2014]
- [83] P. García, J. Roca, A. Varela, C. Nieto, J. Roca, Las redes neuronales artificiales como herramienta para la minería de datos en un contexto financiero. Centro Universitario de la Defensa de San Javier. [En línea]: <<http://repositorio.bib.upct.es/dspace/bitstream/10317/1707/1/rna.pdf>>. [Acceso: 22-Febrero-2014]
- [84] R. Fernández, Técnica de Inteligencia Artificial en Minería de Datos. [En línea]: <<http://pendientedemigracion.ucm.es/info/pslogica/rodolfo2.pdf>>
- [85] S. Pérez, Minería de datos (Redes Neuronales). Universidad Autónoma Metropolitana. [En línea]: < <http://computacion.cs.cinvestav.mx/~sperez/cursos/md/RedesNeuronales.pdf> >. [Acceso: 16-Marzo-2014]
- [86] C. Bravo, S. Maldonado, R. Weber, Seguimiento en Modelos De Regresión Logística. , Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile. [En línea]: < <http://goo.gl/2HJq3h>>. [Acceso: 16-Marzo-2014]
- [87] E. Acuna, Clasificación Supervisada: Regresión Logística. Universidad de Puerto Rico- Mayaguez. [En línea]: <<http://academic.uprm.edu/eacuna/logisacu.pdf>>. [Acceso: 10-Abril-2014]

- [88] C. Soto, Incorporación De Técnicas Multivariantes en un Sistema Gestor de Bases de Datos. Universidad Nacional de Colombia. 2009. [En línea]: <[http://www.bdigital.unal.edu.co/895/1/71335481\\_2009.pdf](http://www.bdigital.unal.edu.co/895/1/71335481_2009.pdf)>. [Acceso: 14-Abril-2014]
- [89] L. Vieira, L. Ortiz, S. Ramírez, Introducción a la Minería de Datos 2012, Edit. E-papers, 2009, p 45-47.
- [90] S. Pérez, J. Puldón, R. Espín , Modelo Clustering para el Análisis en la Ejecución de Procesos de Negocio. Revista Investigación Operacional, 2012, vol. 33, no 33.
- [91] F. Berzal, Clustering. Departamento de Ciencias de la Computación, Universidad de Granada. [En línea]: < <http://elvex.ugr.es/decsai/intelligent/slides/dm/D3%20Clustering.pdf> >. [Acceso: 17-Abril-2014]
- [92] R. Gómez, J. Jiménez, Minería De Datos: Análisis de Datos Electorales. [En línea]: < <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/27.pdf> >. [Acceso: 29-Abril-2014]
- [93] Cliente gráfico: pgAdmin3. [En Línea]<<http://www.dataprix.com/8-cliente-grafico-pgadmin3>>. [Acceso: 29-Abril-2014]
- [94] Webyog, SQLyog MySQL Editor, herramienta GUI Visual. [En Línea]<<https://www.webyog.com/product/sqlyog>>. [Acceso: 29-Abril-2014]
- [95] JenaSoft, Administrador de datos eficiente herramienta de gestión de datos DatAdmin, [En Línea]<<http://www.datadmin.com/features/data-admin-reference-browser>>. [Acceso: 29-Abril-2014]
- [96] SINTES, Bartolomé, PhpMyAdmin, Soluciones. Noviembre 2013 [En Línea]<[http://www.mclubre.org/consultar/webapps/lecciones/webapps\\_phpmyadmin\\_soluciones.html](http://www.mclubre.org/consultar/webapps/lecciones/webapps_phpmyadmin_soluciones.html)>. [Acceso: 11-Mayo-2014]
- [97] Página Oficial Many Eyes. [En Línea] <<http://www-958.ibm.com/software/data/cognos/manyeyes/>>. [Acceso: 23-Mayo-2014]
- [98] S. Heymann, Gephi, [En Línea] < <http://sebastien.pro/gephi-esnam.pdf>>. [Acceso: 25-Mayo-2014]
- [99] E. Botta, J. Cabrera, Minería de textos: una herramienta útil para mejorar la gestión del bibliotecario en el entorno digital. 2007 [En Línea] <[http://bvs.sld.cu/revistas/aci/vol16\\_4\\_07/aci051007.html](http://bvs.sld.cu/revistas/aci/vol16_4_07/aci051007.html)>. [Acceso: 25-Mayo-2014]
- [100] Usos de Data Mining y KDD en Marketing Research. Facultad Ciencias Exactas y Naturales UB. Noviembre 2008. [En Línea]<<http://www.alephzero.com.ar/library/charla.pdf>>. [Acceso: 25-Mayo-2014]

- [101] N. TEDESCHI, Web Services, un ejemplo práctico. [En Línea]<<http://msdn.microsoft.com/es-es/library/bb972248.aspx#mainSection>>. [Acceso: 25-Mayo-2014]
- [102] EcuRed, Archivos de procesamiento por lotes .BATC. [En Línea]<[http://www.ecured.cu/index.php/Archivos\\_de\\_procesamiento\\_por\\_lotes\\_.BATC](http://www.ecured.cu/index.php/Archivos_de_procesamiento_por_lotes_.BATC)>
- [103] El sistema operativo GNU. [En Línea]<<https://www.gnu.org/licenses/licenses.es.html>>. [Acceso: 23-Mayo-2014]
- [104] J. Candás, Minería de datos: conceptualización y Utilidad. Septiembre 2006. [En Línea]<<http://bid.ub.edu/17canda2.htm>>. [Acceso: 23-Mayo-2014]
- [105] Sistemas gestores de bases de datos. [En Línea]<<ftp://ftp.puce.edu.ec/Facultades/Ingenieria/Sistemas/Base%20de%20Datos%20II/Sistemas%20Gestores%20de%20Bases%20de%20Datos%20Capitulo%201.pdf>>. [Acceso: 23-Mayo-2014]
- [106] G. Moreno, Técnicas Más Usadas en la Minería De Datos. Octubre 2007. [En Línea]<<http://gamoreno.wordpress.com/2007/10/03/tecnicas-mas-usadas-en-la-mineria-de-datos/>>. [Acceso: 23-Mayo-2014]
- [107] M. SILVA, Minería de Datos y Descubrimiento del Conocimiento. [En Línea]<[http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria\\_de\\_Datos\\_y\\_KDD.pdf](http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_de_Datos_y_KDD.pdf)>. [Acceso: 27-Mayo-2014]
- [108] Gestión Del Conocimiento. [En Línea]<[http://www.fundibeq.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/gestion\\_del\\_conocimiento.pdf](http://www.fundibeq.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/gestion_del_conocimiento.pdf)>. [Acceso: 27-Mayo-2014]
- [109] J. Orallo, C. Ramírez, Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software. [En Línea]<<http://users.dsic.upv.es/~jorallo/docent/doctorat/t2a.pdf>> [Acceso: 27-Mayo-2014]
- [110] P. Britos, V. García. Propuesta de Procesos de Explotación de Información. En XV Congreso Argentino de Ciencias de la Computación. 2009.

## **k. Anexos**

### **ANEXO 1: Acuerdo de Confidencialidad**

#### **DECLARACIÓN DE CONFIDENCIALIDAD**

Janina Silvana Ortiz Pasaca con cédula de identidad Nro 1104902612 (en adelante la declarante). Declara lo siguiente:

#### **PRIMERO: Antecedentes**

- 1) La declarante va a participar en el proyecto fin de carrera “Análisis del proceso de Evaluación al Desempeño Docente, aplicando Técnicas de Minería de Datos”, dirigido por el Ing. Luis Antonio Chamba Eras, en calidad de director del proyecto.
- 2) Por el presente documento se regula el tratamiento que la declarante ha de dar a la información a la que pueda tener acceso en el desarrollo de las tareas de investigación que se realicen en dicho proyecto, el cual se regulará por las disposiciones contenidas en las siguientes cláusulas.

#### **SEGUNDO: Información Confidencial**

La información requerida a datos de evaluación, métodos y resultados científicos o técnicos utilizados u obtenidos durante la realización del proyecto de investigación o una vez realizado el mismo, se considerará siempre Información Confidencial.

#### **TERCERO: Excepciones**

No será considerada como información Confidencial:

- a) La información que la declarante pueda probar que tenían en su legítima posesión con anterioridad al conocimiento de la Información Confidencial.
- b) La información que la declarante pueda probar que era de dominio público en la fecha de la divulgación o pase a serlo, con posterioridad, por haberse publicado o por otro medio, sin intervención ni negligencia de la declarante.
- c) La información que la declarante pueda probar que corresponde en esencia a información facilitada por terceros, sin restricción alguna sobre su divulgación, en virtud de un derecho de la declarante a recibirla.

**CUARTO: Secreto de la información Confidencial**

La declarante se compromete a mantener totalmente en secreto la Información Confidencial recibida en relación con el proyecto referido anteriormente y no divulgarla a terceros durante la vigencia de esta Declaración de Confidencialidad.

Además y a través de la presente acta de confidencialidad se da por sentado que no existe peligro alguno en cuanto a los datos prestados hacia la declarante, ya que los mismos se usarán solo para los fines especificados anteriormente que es hacia el proyecto fin de carrera.

Asimismo, la declarante se compromete a emplear la Información Confidencial, exclusivamente, en el desempeño de las tareas que tenga encomendadas en dicho proyecto.

**QUINTO: Duración**

La obligación de la declarante respecto al mantenimiento del compromiso de secreto de la Información Confidencial, será indefinido para fines de investigación a partir de la recepción de la Información Confidencial.

Loja 17 de Enero del 2014.

Janina Silvana Ortiz Pasaca.



## ANEXO 2: Autorización Acceso a los Datos de Evaluación de Desempeño Docente



**UNIVERSIDAD NACIONAL DE LOJA**  
**UNIDAD DE TELECOMUNICACIONES E INFORMACIÓN**

*Of. N.009-UTI-UNL*  
*Loja, 07 de enero de 2014*

*Srta.*  
*Janina Silvana Ortiz Pasaca*  
*TESISTA DE LA CARRERA DE INGENIERIA EN SISTEMAS*  
*Ciudad.-*

*De mi Consideración.*

*Mediante la presente me permito saludar a su persona para hacerle la entrega de copia de la Base de Datos del Sistema de Evolución del Docente el mismo que va en un CD. Para realizar el tema de Tesis "Análisis del Proceso de Evaluación al Desempeño Docente aplicando técnica de Minería de Datos."*

*Particular que le comunico para fines pertinentes.*

*Atentamente.*

  
*Ing. Milton Labanda*  
*RESPONSABLE DEL SGA.*

  
*07/01/2014*

## ANEXO 3: Autorización Acceso a los Datos del Web Services



**UNIVERSIDAD NACIONAL DE LOJA**  
**UNIDAD DE TELECOMUNICACIONES E INFORMACIÓN**

*Of. N.263-UTI-UNL*  
*Loja, 16 de abril de 2014*

*Sta.*  
*Janina Silvana Ortiz*  
*Ciudad.*

*De mi Consideración:*

*Mediante la presente me dirijo a su persona para comunicarle que se hace la entrega de la clave y el usuario para el acceso de datos a través de los Web Services del Sistema*

*Usuario.-*

*Clave.-*

*Particular que pongo a su conocimiento para los fines pertinentes.*

*Atentamente,*

*Ing. Milton Palacios Morocho*  
**DIRECTOR DE LA UNIDAD**  
**DE TELECOMUNICACIONES E INFORMACIÓN**

## ANEXO 4: Cuestionario de Evaluación Docente y Satisfacción Estudiantil.

### Cuestionario Evaluación Docente



UNIVERSIDAD NACIONAL DE LOJA  
Comisión de Evaluación Interna

#### CUESTIONARIO

Para evaluar a su docente/a dígnese marcar una X en el casillero de su preferencia, utilice la siguiente escala:

- 4 = Totalmente de acuerdo
- 3 = Mayoritariamente de acuerdo
- 2 = Minoritariamente de acuerdo
- 1 = En desacuerdo

#### CAPACIDAD PROFESIONAL (dominio científico técnico)

CPF.1.03	1. El docente demuestra dominio científico de los contenidos que contribuyen a explicar la problemática del módulo ( <b>trata los contenidos en toda su complejidad, los relaciona, aclara y ejemplifica</b> ).	4	3	2	1
CPF.1.04	1. El docente trabaja con contenidos pertinentes o actualizados ( <b>relacionados con los más recientes avances del conocimiento científico técnico de la profesión</b> ).	4	3	2	1
	2. Los contenidos tratados satisfacen las expectativas estudiantiles.	4	3	2	1
CPF.1.05	1. Los contenidos seleccionados por el docente corresponden al nivel de formación del estudiante (nivel profesional) y son tratados con profundidad	4	3	2	1
CPF.1.06	1. El docente señala bibliografía básica y complementaria pertinente a los objetivos de la asignatura curso o taller.	4	3	2	1
CPF.1.07	1. El docente orienta y apoya al estudiante en actividades de investigación formativa (lectura comprensiva, manejo de fuentes de consulta, identificación y planteamiento de problemas, manejo y procesamiento de datos, elaboración de conclusiones).	4	3	2	1
¿Qué sugiere Ud., para que el docente mejore su capacidad profesional?					

#### CAPACIDAD PEDAGÓGICA (desarrollo del proceso enseñanza-aprendizaje)

CPG.2.01	1. El docente se comunica fácilmente con sus estudiantes ( <b>dialoga, presta atención, escucha, ayuda a que los estudiantes se sientan</b>	4	3	2	1
----------	---	---	---	---	---

	<b>valorados, etc.).</b>				
CPG.2.02	1. El docente al inicio del tratamiento del sílabo socializa con los estudiantes la programación respectiva y establece las responsabilidades de cada uno (encuadre).	4	3	2	1
CPG.2.03	1. El docente lleva a cabo un proceso lógico, ordenado y claro en el trabajo de aula y en otros escenarios.	4	3	2	1
	2. El docente brinda asesoría y apoyo para fortalecer los aprendizajes ( <b>organiza el trabajo en los grupos, orienta las actividades a desarrollar, amplía la información, señala las bases de datos a las que tiene acceso la Universidad</b> ).	4	3	2	1
CPG.2.04	1. El docente revisa periódicamente el trabajo individual y grupal que realizan los estudiantes dentro y fuera del aula.	4	3	2	1
	2. El docente asegura a través de procesos permanentes de evaluación que los estudiantes logren los aprendizajes previstos en el sílabo y utiliza diversos instrumentos de evaluación (pruebas orales, escritas, trabajos prácticos, tareas de resolución individual, de construcción grupal).	4	3	2	1
	3. El docente estimula la participación activa de los estudiantes y controla sean honestos en la realización de las tareas.	4	3	2	1
	4. El docente manifiesta entusiasmo y buen humor en el desarrollo de la clase				
CPG.2.05	1. El docente revisa los avances de los trabajos y los devuelve oportunamente con observaciones y recomendaciones.	4	3	2	1
CPG.2.06	1. El docente lleva un registro de las actividades académicas cumplidas por los estudiantes y les informa periódicamente sobre su situación (cumplimiento de tareas, consultas, exposiciones, reportes, lecciones, informes).	4	3	2	1
	2. El docente comunica y analiza con los estudiantes los resultados de la evaluación de los aprendizajes y da la retroalimentación correspondiente.	4	3	2	1
CPG.2.07	1. El docente planifica, supervisa y evalúa las prácticas académicas de los estudiantes ( <b>prácticas necesarias para desarrollar las habilidades y destrezas que demanda el desempeño de las prácticas profesionales</b> ).	4	3	2	1
	2. El nivel de exigencia del docente en el cumplimiento de tareas es el adecuado ( <b>ajustado a las necesidades de formación profesional de los estudiantes</b> ).	4	3	2	1
CPG.2.08	1. El docente propicia siempre un ambiente adecuado de trabajo en el aula para que los estudiantes aprendan (ambiente de confianza, respeto y atención a los estudiantes).	4	3	2	1
	2. El docente motiva y hace su trabajo dinámico e interesante.	4	3	2	1
CPG.2.09	1. El docente utiliza gabinetes, laboratorios, áreas experimentales, equipos, talleres u otros escenarios en donde los estudiantes puedan vincular la teoría con la práctica ( <b>de acuerdo a lo planificado en el sílabo</b> ).	4	3	2	1
CPG.2.10	1. El docente cumple con los contenidos programados en el sílabo.	4	3	2	1
CPG.2.11	1. El docente es claro en sus exposiciones y explicaciones (se hace entender con facilidad).	4	3	2	1
	2. El docente resuelve las dudas y problemas planteados por el estudiante.	4	3	2	1
	3. El docente logra aprendizajes significativos en los estudiantes ( <b>aprendizajes útiles para resolver problemas relacionados con</b>	4	3	2	1

	<b>las prácticas profesionales).</b>				
CPG.2.12	1. El docente articula acciones investigativas en el desarrollo de la asignatura (consultas, procesos de análisis, de síntesis, observaciones, elaboración de informes y conclusiones).	4	3	2	1
CPG.2.13	1. El docente incorpora las nuevas tecnologías de la información y la comunicación en el proceso enseñanza aprendizaje, para mejorar los resultados de aprendizaje del estudiante (se comunica con el estudiante a través de chats, blogs, wikis, pag. web, plataforma virtual, e-mail, video).				
	2. El docente utiliza la técnica del trabajo autónomo (TAE) para la realización de algunas actividades planificadas en el sílabo y especifica con claridad las formas de interacción con los estudiantes para darles el acompañamiento adecuado.	4	3	2	1
CPG.2.15	1. El docente elabora documentos de apoyo para ampliar el conocimiento de los contenidos de la asignatura, curso o taller y los entrega a los estudiantes.	4	3	2	1
	1. El docente elabora material didáctico para fortalecer el aprendizaje de los estudiantes.	4	3	2	1
CPG.2.16	1. La relación entre el docente y el alumno es cordial.	4	3	2	1
CPG.2.17	1. El docente promueve en el aula y fuera de ella espacios de discusión, análisis y debate sobre temáticas relacionadas con la problemática de la profesión que se aborda en la asignatura, curso, o taller.	4	3	2	1
	2. El docente propicia aprendizajes en los estudiantes, a través de actividades de investigación formativa (los estudiantes aprenden a identificar, definir y plantear problemas; a buscar, manejar y procesar la información; a elaborar análisis y síntesis; a elaborar instrumentos para investigar; a elaborar conclusiones, informes y a trabajar en equipo).				
¿Qué sugiere para que mejore su capacidad pedagógica?					

### PRÁCTICA DE VALORES

PV.3.01	1. El docente es ejemplo de honestidad, tolerancia y de respeto.	4	3	2	1
	2. El docente tiene un trato igual para con todos los estudiantes.	4	3	2	1
	3. El docente demuestra responsabilidad en su trabajo.	4	3	2	1
	4. El docente demuestra una actitud solidaria con los estudiantes ante situaciones de calamidad y de logros alcanzados.	4	3	2	1
	5. El docente demuestra lealtad y compromiso con la institución.	4	3	2	1
PV.3.02	1. El docente propicia un ambiente favorable para que los estudiantes conozcan sus derechos y responsabilidades y aprendan a ejercerlos.	4	3	2	1
	2. Su trabajo en el aula contribuye efectivamente a la formación integral del estudiante.	4	3	2	1
PV.3.03	1. El docente evalúa, acredita y califica con justicia, equidad y transparencia.	4	3	2	1
PV.3.04	1. El docente asiste normalmente a su trabajo.	4	3	2	1
	2. El docente cumple con responsabilidad las horas y calendarios de trabajo establecidos institucionalmente.	4	3	2	1
PV.3.05	1. El docente cumple con el 100% de lo planificado.	4	3	2	1
¿Qué sugiere para que mejore su práctica de valores?					

## Cuestionario Satisfacción Estudiantil



**UNIVERSIDAD NACIONAL DE LOJA**  
*Comisión de Evaluación Interna*

### CUESTIONARIO

Para evaluar a su docente/a dígnese marcar una X en el casillero de su preferencia, utilice la siguiente escala:

- 4 = Totalmente de acuerdo
- 3 = Mayoritariamente de acuerdo
- 2 = Minoritariamente de acuerdo
- 1 = En desacuerdo

**"Mencione en qué medida se siente satisfecho (a) con:"**

<b>Proceso Docente"</b>
<i>La planificación del módulo, unidad, curso o taller</i>
<i>El contenido teórico del módulo, unidad, curso o taller.</i>
<i>Los materiales impresos y/o material didáctico disponible.</i>
<i>La forma como el docente planifica y organiza las clases.</i>
<i>El dominio de los contenidos y actualización del docente.</i>
<i>El proceso de comunicación y participación de los estudiantes que el docente promueve en el aula (motivación para el aprendizaje).</i>
<i>La relación docente – estudiantes en el aula y fuera de ella (el docente es cordial, genera confianza, sabe escuchar, crea un ambiente favorable de trabajo).</i>
<i>La valoración de los saberes ancestrales y el respeto a la diversidad cultural que fomenta el docente.</i>
<i>La metodología de enseñanza utilizada por el docente.</i>
<i>La relación teoría – práctica que el docente realiza en el tratamiento de los contenidos del módulo, unidad, curso o taller (orientada hacia la práctica profesional).</i>
<i>El nivel de exigencia y rigor científico promovido por el docente.</i>
<i>El interés que pone y el esfuerzo que realiza el docente para que los estudiantes aprendan.</i>
<i>Las actividades extra- clase que le envía el docente para reforzar los aprendizajes y optimizar su formación.</i>
<i>Los mecanismos que utiliza el docente para evaluar los aprendizajes de los estudiantes (tareas, lecciones, reportes, pruebas oral y escrita, prácticas).</i>
<i>La retroalimentación que da el docente después de la evaluación (trabaja sobre las deficiencias encontradas, analiza, aclara y explica).</i>
<i>La asesoría del profesor en espacios intra- aula o cuando usted requiere de su apoyo."</i>
<i>El apoyo de la formación que recibe en el módulo, unidad, curso o taller para sus prácticas profesionales.</i>
<i>Las nuevas tecnologías de la Comunicación y la información que su docente incorpora a la enseñanza, en apoyo al trabajo autónomo de los estudiantes (e-mail, blogs, wikis, chats, plataforma virtual).</i>

<i>Las estrategias de investigación formativa que implementa el docente para fomentar el espíritu investigativo y creatividad de los estudiantes (lecturas de la realidad, manejo de fuentes bibliográficas, de datos, confrontación de ideas, estudio de casos, análisis de situaciones problemáticas, recolección de información, etc.).</i>
<i>La bibliografía que señala el docente.</i>

**Mencione en qué medida se siente satisfecho (a) con:**

<b>Logros Alcanzados</b>
<i>Sus conocimientos teóricos adquiridos.</i>
<i>Sus aprendizajes logrados y que los puede demostrar.</i>
<i>Las calificaciones asignadas por el docente</i>
<i>El trabajo intelectual que ha logrado desarrollar (hace análisis, elabora síntesis, razonamientos, establece relaciones, comparaciones)</i>
<i>Los nuevos valores integrados a su formación</i>
<i>Su capacidad desarrollada para trabajar en equipo (se prepara, participa, acepta sugerencias, llega a acuerdos, respeta los puntos de vista de los demás, es puntual)</i>
<i>Su capacidad para resolver problemas y casos de la especialidad</i>
<i>Sus habilidades y destrezas desarrolladas, en relación con las prácticas profesionales</i>
<i>Su capacidad para actuar con autonomía e iniciativa</i>
<i>El desarrollo de una actitud positiva hacia el cambio e innovación</i>
<i>Su capacidad de argumentación y reflexión crítica desarrollada</i>

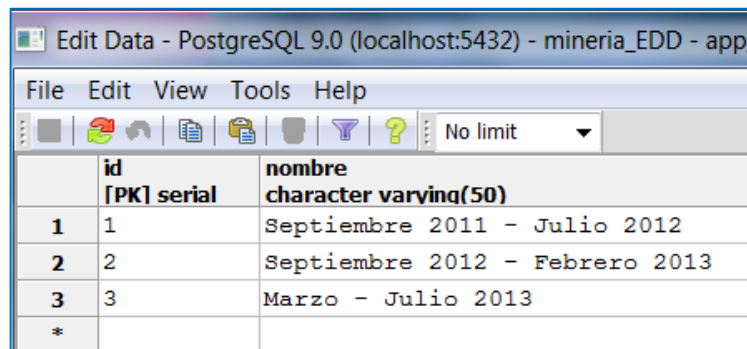
**"Mencione en qué medida se siente satisfecho (a) con:**

<b>Comportamiento Ético del Docente</b>
<i>El tratamiento honesto y respetuoso del docente</i>
<i>La responsabilidad que demuestra el docente en su trabajo.</i>
<i>El cumplimiento del docente con los horarios y calendarios de trabajo establecidos institucionalmente</i>
<i>La firmeza del docente en las decisiones tomadas y para que los estudiantes respeten los acuerdos</i>
<i>La equidad, justicia, transparencia y oportunidad con que el docente evalúa, acredita y califica el desempeño de los estudiantes</i>
<i>El cumplimiento de lo planificado en el módulo, unidad, curso o taller</i>

<b>Otros Aspectos</b>
<i>¿Qué otros aspectos que no se hayan considerado en el Cuestionario le producen Satisfacción?</i>
<i>a. Respecto al proceso docente</i>
<i>b. Respecto a los logros alcanzados</i>
<i>c. Respecto al comportamiento ético del docente</i>

## ANEXO 5: Datos Recopilados

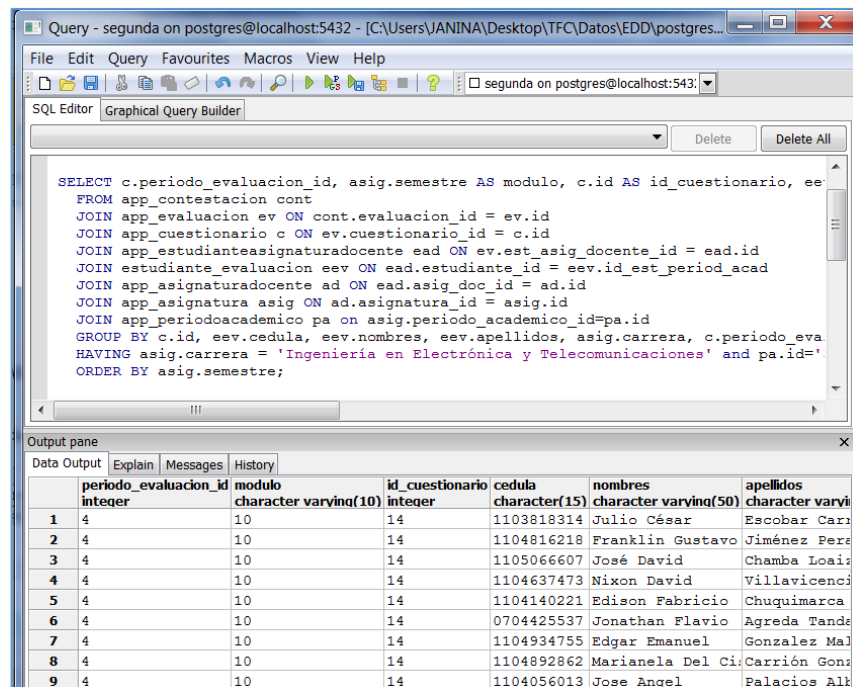
Para la descripción de los datos se hizo uso de consultas dentro de la base de datos que permitan la exploración y por ende la preparación de los datos, de tal manera que se pudo relacionar con los mismos, conocer la estructura como el formato de los datos disponibles. La siguiente figura presenta los periodos académicos en los cuales se llevó a cabo el proceso de evaluación (Ver Figura 1).



	id [PK] serial	nombre character varying(50)
1	1	Septiembre 2011 - Julio 2012
2	2	Septiembre 2012 - Febrero 2013
3	3	Marzo - Julio 2013
*		

Figura 1: Periodos académicos en cuanto a la EDD

De la misma forma se pudo obtener el número de estudiantes que evaluaron a sus docentes dentro de estos periodos, por carreras y módulos (Ver Figura 2).



```

SELECT c.periodo_evaluacion_id, asig.semestre AS modulo, c.id AS id_cuestionario, ee
FROM app_contestacion cont
JOIN app_evaluacion ev ON cont.evaluacion_id = ev.id
JOIN app_cuestionario c ON ev.cuestionario_id = c.id
JOIN app_estudianteasignaturadocente ead ON ev.est_asig_docente_id = ead.id
JOIN estudiante_evaluacion eev ON ead.estudiante_id = eev.id_est_period_acad
JOIN app_asignaturadocente ad ON ead.asig_doc_id = ad.id
JOIN app_asignatura asig ON ad.asignatura_id = asig.id
JOIN app_periodoacademico pa ON asig.periodo_academico_id=pa.id
GROUP BY c.id, eev.cedula, eev.nombres, eev.apellidos, asig.carrera, c.periodo_eva
HAVING asig.carrera = 'Ingeniería en Electrónica y Telecomunicaciones' and pa.id='
ORDER BY asig.semestre;

```

periodo_evaluacion_id	modulo	id_cuestionario	cedula	nombres	apellidos
integer	character varying(10)	integer	character(15)	character varying(50)	character varyi
1	10	14	1103818314	Julio César	Escobar Carr
2	10	14	1104816218	Franklin Gustavo	Jiménez Peré
3	10	14	1105066607	José David	Chamba Loai
4	10	14	1104637473	Nixon David	Villavicenci
5	10	14	1104140221	Edison Fabricio	Chuquimarca
6	10	14	0704425537	Jonathan Flavio	Agreda Tande
7	10	14	1104934755	Edgar Emanuel	Gonzalez Mal
8	10	14	1104892862	Marianela Del Ci	Carrión Gonz
9	10	14	1104056013	Jose Angel	Palacios Alk

Figura 2: Estudiantes del AEIRNNR



A continuación se presenta la generación de la consulta para obtener los docentes por periodo académico (Ver Figura 3)

The screenshot shows a PostgreSQL query editor window titled "Query - mineria\_edd en postgres@localhost:5432". The SQL query is as follows:

```
SELECT de.cedula, de.nombres, de.apellidos, asig.carrera, pa.id AS periodo_Academico
FROM app_contestacion cont
JOIN app_evaluacion ev ON cont.evaluacion_id = ev.id
JOIN app_cuestionario c ON ev.cuestionario_id = c.id
JOIN app_estudianteasignaturadocente ead ON ev.est_asig_docente_id = ead.id
JOIN estudiante_evaluacion eev ON ead.estudiante_id = eev.id_est_period_acad
```

The results are displayed in a table with the following columns: cedula, nombres, apellidos, carrera, and periodo\_academico. The table contains 11 rows of data.

	cedula character varying(15)	nombres character varying(100)	apellidos character varying(100)	carrera character varying(100)	periodo_academico integer
1	1104179559	Juan Pablo	Cabrera Samaniego	Ingeniería en Elect	3
2	1102293402	Klever Filiberto	Carrión Gordillo	Ingeniería en Elect	3
3	1100613460	Luis Alberto	Cuenca Cuenca	Ingeniería en Elect	3
4	1102103247	Mercy Edilda	Feijoo Cisneros	Ingeniería en Elect	3
5	1103532766	Juan Manuel	Galindo Vera	Ingeniería en Elect	3
6	1104199086	Rodolfo Pabel	Merino Vivanco	Ingeniería en Elect	3
7	1103778633	Diego Vinicio	Orellana Villavicen	Ingeniería en Elect	3
8	1102734793	Luisa Inés	Sinche Freire	Ingeniería en Elect	3
9	1103401475	Juan Carlos	Solano	Ingeniería en Elect	3
10	1103937395	Marcelo Fernando	Valdiviezo Condolo	Ingeniería en Elect	3
11	1103872360	Ivan Rodrigo	Velepucha Morillo	Ingeniería en Elect	3

Figura 3: Docentes por Periodo Académico

En la siguiente se presenta las carreras objeto de estudio, ofertadas dentro del (Ver Figura 4)

The screenshot shows a PostgreSQL query editor window displaying a table with the following columns: id and carrera. The table contains 6 rows of data.

	id [PK] integer	carrera character varying(255)
1	1	Ingeniería en Sistemas AEIRNNR
2	32	Ingeniería en Electromecánica AEIRNNR
3	33	Ingeniería en Geología Ambiental y Ordenamiento Territorial AEIRNNR
4	34	Ingeniería en Electrónica y Telecomunicaciones AEIRNNR
5	35	Ingeniería en Geología Ambiental y Ordenamiento Territorial (especial) AEIRNNR
6	36	Tecnología en Electricidad y Control Industrial AEIRNNR

Figura 4: Carreras AEIRNNR

## ANEXO 6: Herramientas útiles en el proceso de Minería de Datos

A continuación se presentan cada una de las herramientas tomadas para el análisis, dando énfasis a sus características. (Ver Figura 1).

- **Knime**

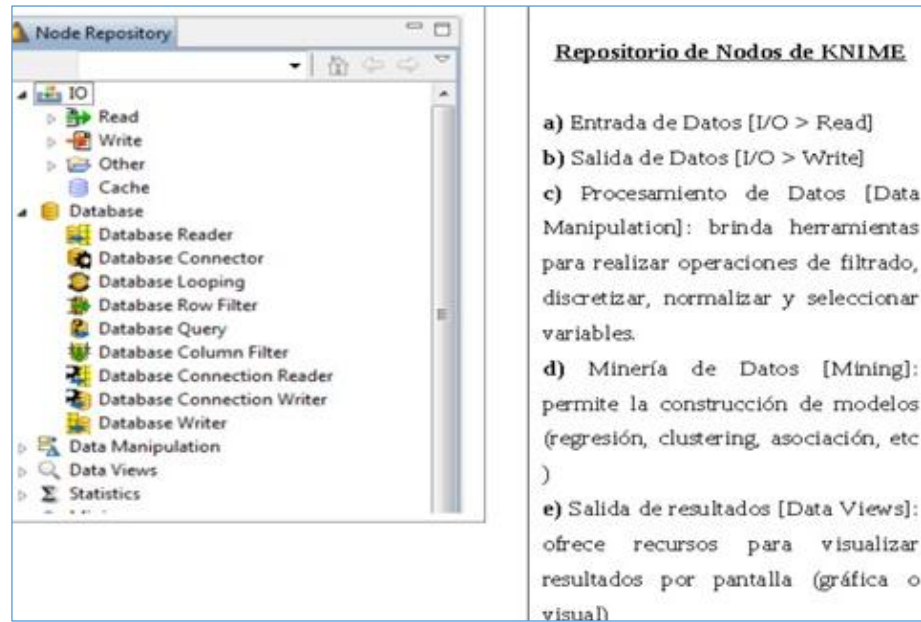


Figura 1: Interfaz de Knime (nodos).

La herramienta Knime presenta la posibilidad de presentar los datos por medio de histogramas (Ver Figura 2)

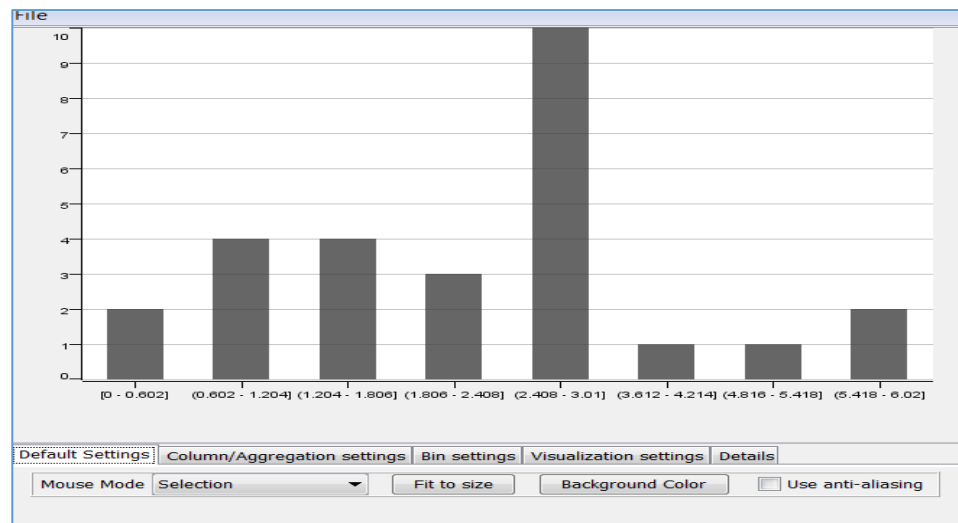


Figura 2: Visualización de histogramas en Knime

A más de ello permite la creación de modelos estadísticos y de minería de datos, como árboles de decisión, regresiones por medio de sus operadores (Ver Figura 3)

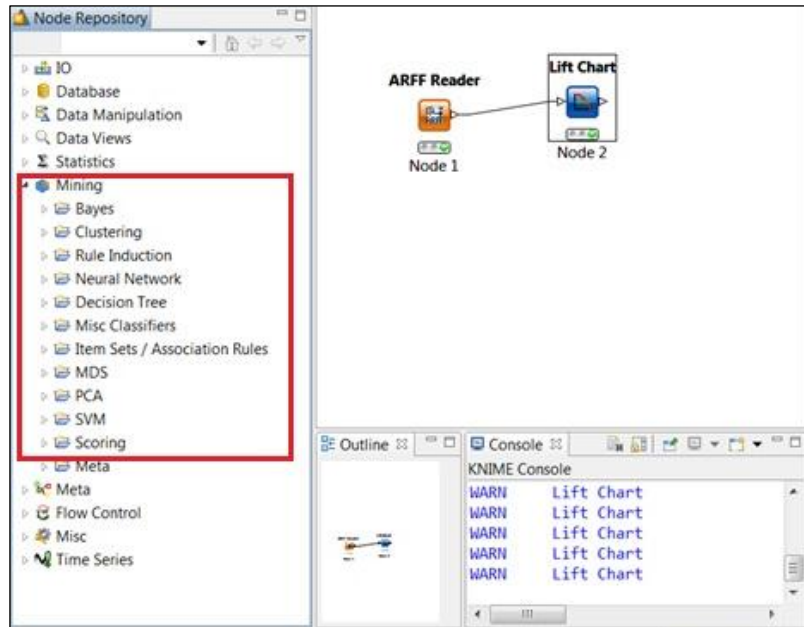


Figura 3: Técnicas de minería de datos y modelos estadísticos en Knime

La visualización de Curvas Roc también es posible en el análisis con Knime (Ver Figura 4)

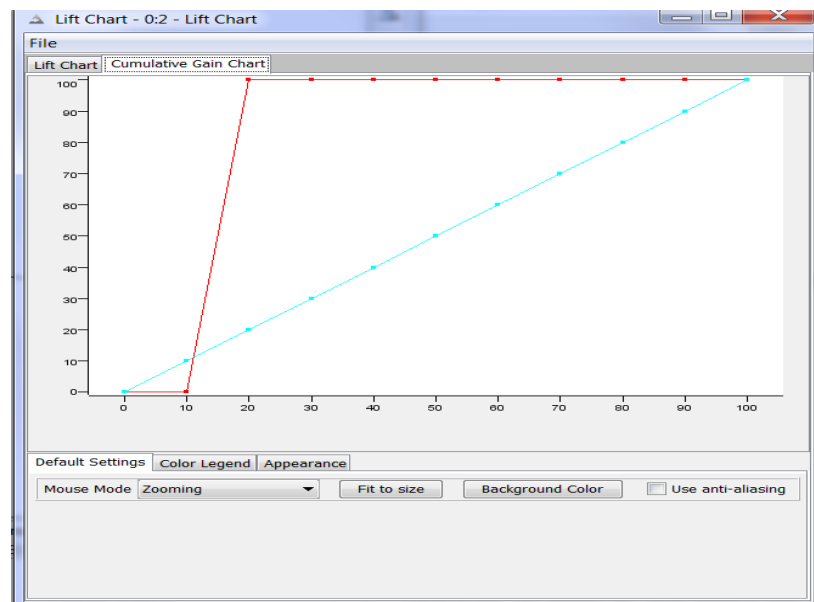


Figura 4: Curvas ROC.

- **Weka**

La interfaz de weka cuenta con cuatro escenarios en los cuales se puede trabajar (ver figura 5-9)

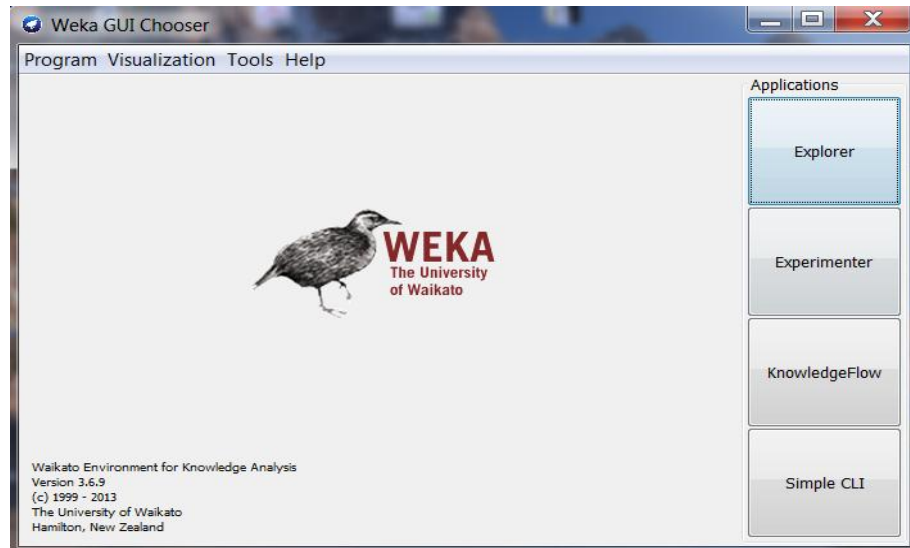


Figura 5: Escenarios de trabajo Weka

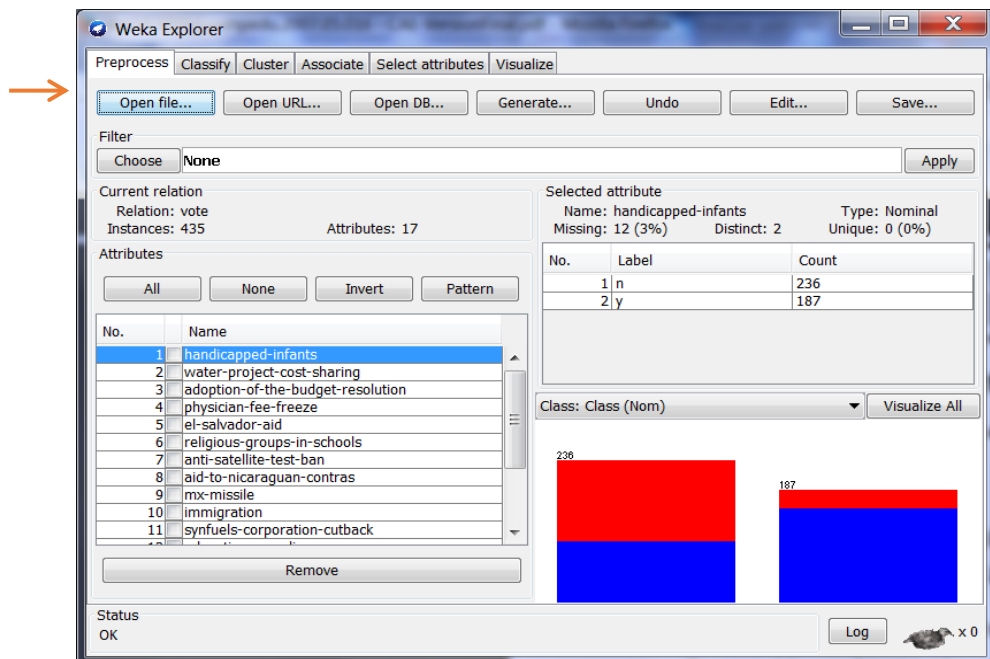


Figura 6: Escenario Explorer: técnicas para el pre procesamiento de datos y modelado.

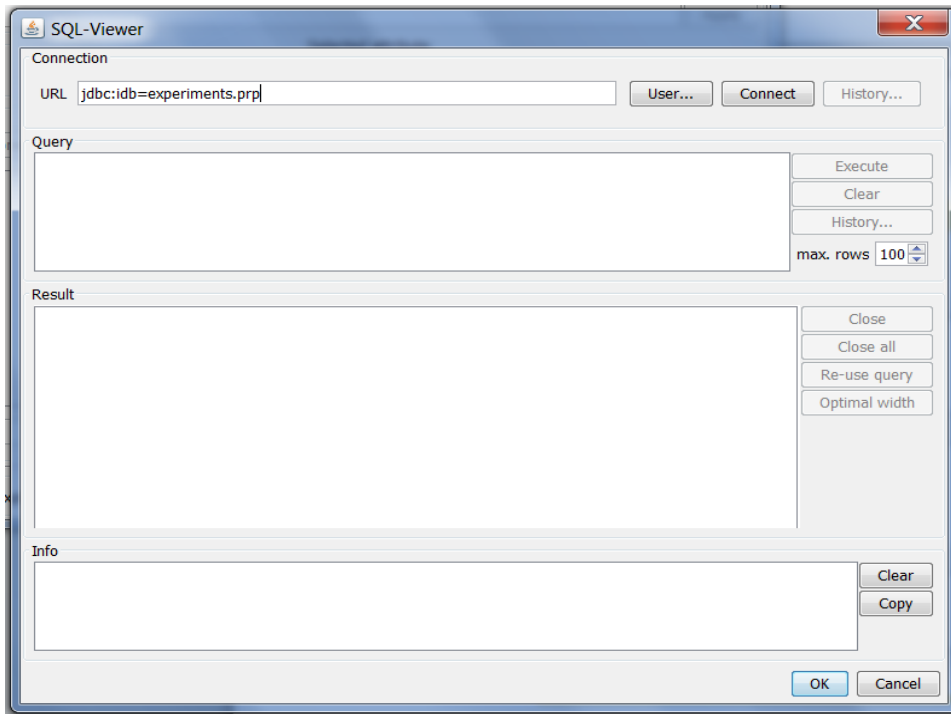


Figura 7: Escenario Explorer: conexión JDBC.

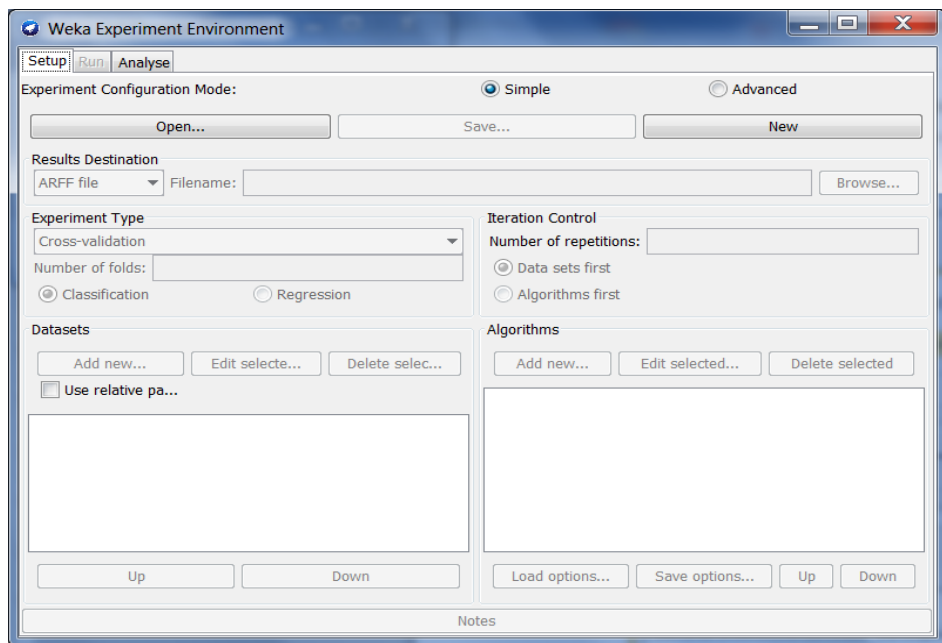


Figura 8: Escenario Experiment.

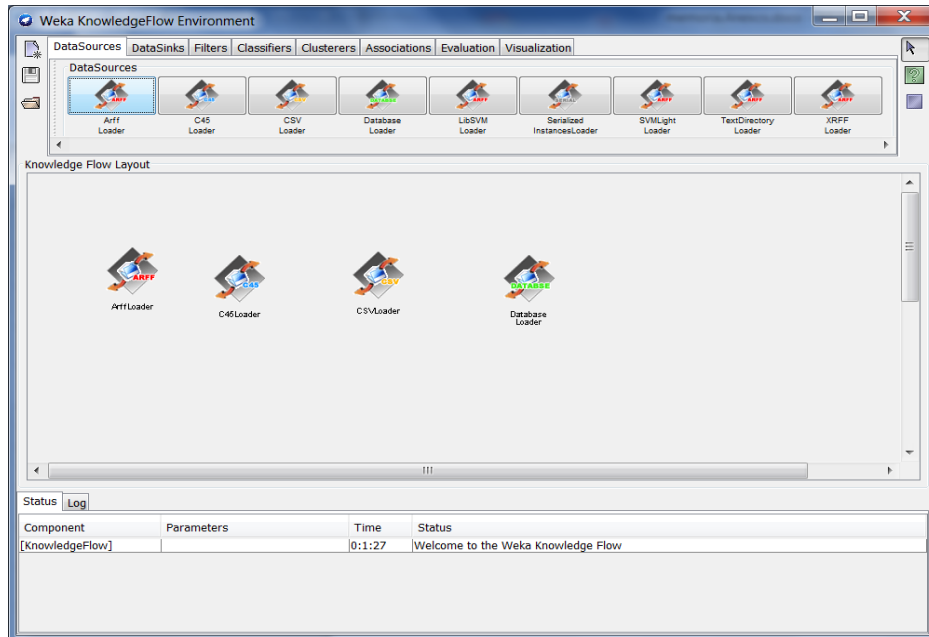


Figura 9: Escenario KnowledgeFlow: flujos de datos en Weka.

- **Rapidminer**

Una de las ventajas que presenta rapidminer como se mencionó anteriormente es la extensa gama de operadores que ofrece, además de poder integrar operadores de weka en la generación del modelo (Ver Figura 10)

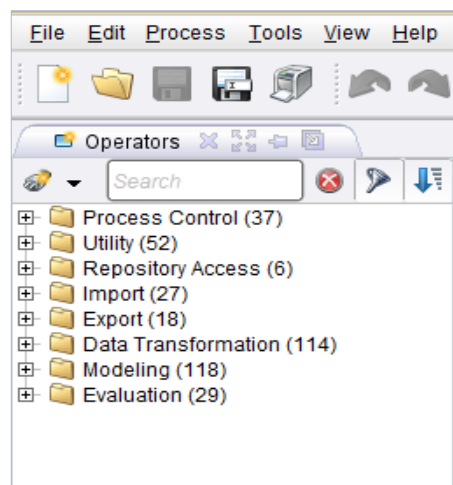


Figura 10: Operadores de RapidMiner.

Rapidminer también permite la jerarquización de operadores, el análisis gráfico, cuenta algoritmos que se aplican en la solución y generación de cualquier modelo contemplado dentro de la minería (Ver Figura 11, 12)

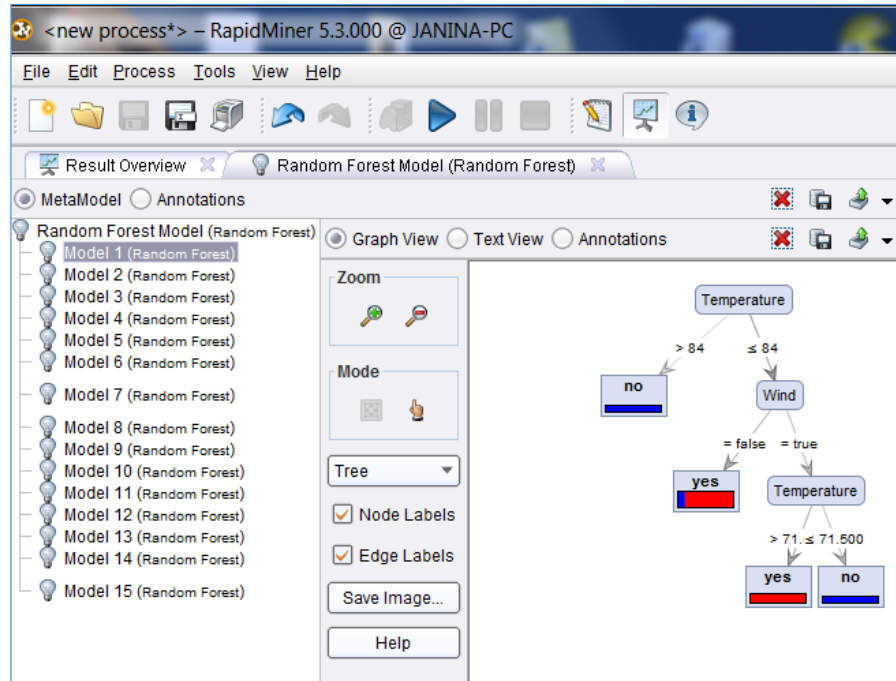


Figura 11: Gráficos y herramientas de visualización de datos

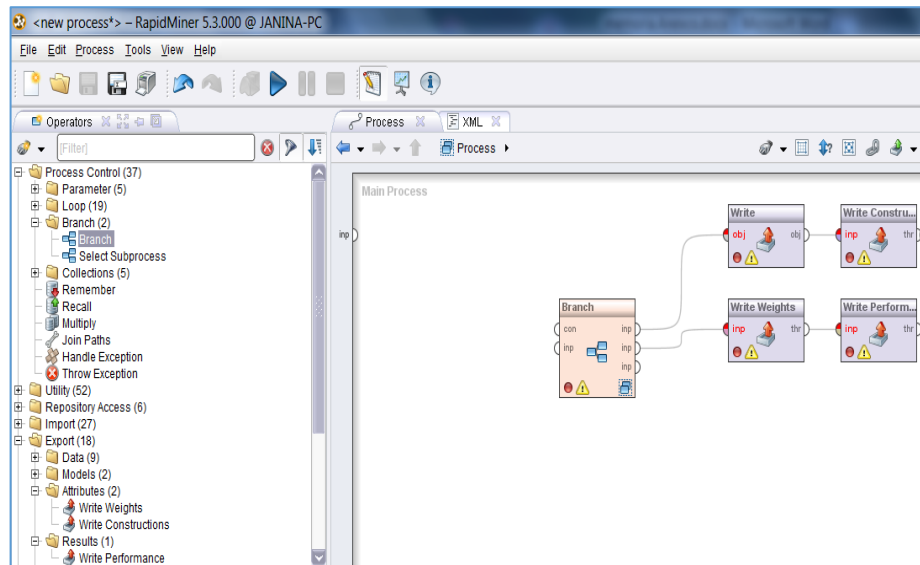


Figura 12: Jerarquización de Operadores.

- **Many Eyes**

Many eyes es una herramienta de apoyo al proceso de Minería de Texto, útil para extraer conocimiento o comportamientos aplicados hacia un texto plano (ver Figura 13)



Figura 13: Herramienta Many Eyes.

La herramienta permite un análisis gráfico en cuanto a textos, libros, revistas, discursos, o todo cuanto posea palabras de las cuales se requiera obtener un conocimiento determinado.

- **Gephi**

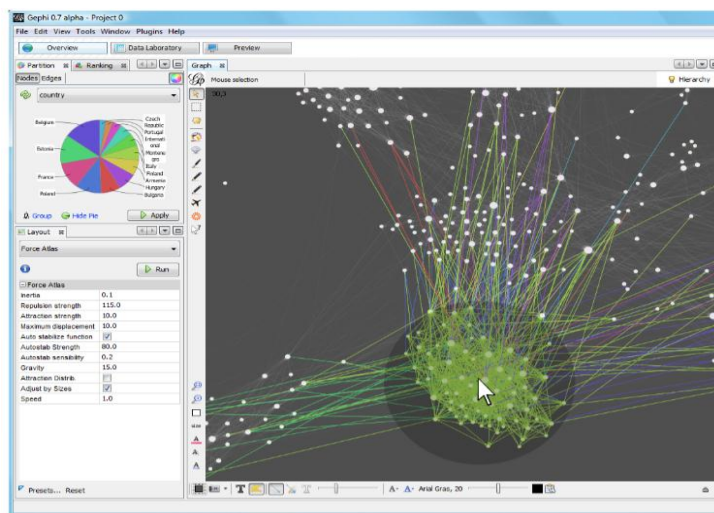


Figura 14: Herramienta Gephi



Gephi al igual que Many Eyes es una Herramienta muy poderosa en cuanto a análisis y en los últimos tiempos también ha sido tomada para la realización de diferentes trabajos investigativos con minería de texto

- **Textalyser**

Textalyser también es una herramienta potente en el análisis con minería de texto, presenta buenas características y es de uso libre (Ver Figura 15).

**Textalyser**

Bienvenido a la herramienta de análisis de textos en línea, las estadísticas detalladas de su texto, ideal para traductores (citando), para los webmasters (ranking) o para los usuarios normales, a conocer el tema de un texto. Ahora, con nuevas características como el análisis de los grupos de palabras, encontrar la densidad de palabras clave, analizar la importancia de la palabra o expresiones. Los webmasters pueden analizar los enlaces de sus páginas. Más instrucciones están a punto de ser escrito, por favor envíenos sus comentarios!

Introduzca el texto a analizar aquí:

o analizar un sitio web:

o seleccione el archivo desde el disco duro local:  
 Ningún archivo seleccionado

Opciones de análisis:

Caracteres mínimo por palabra:

Especial palabra o expresión para analizar:

Número de palabras a analizar:

No haga caso de los números:

Entrar la consulta (sólo para los sitios web):

Aplicar stoplist:

Aplicar propia lista de palabras irrelevantes (Separe con espacios en blanco):

Hacer un análisis de enlaces:

Frases polyword exhaustivos:

Figura 15: Interfaz Herramienta Textalyser

## **ANEXO 7: Procesos y pasos de Minería de Datos operadores en Rapid Miner.**

Los procesos que se emplean dentro de la presente investigación, ya sea en el diseño, generación y evaluación de modelos, requiere el empleo de operadores disponibles dentro de rapid-miner. Dichos operadores forman secuencias de ejecución de procesos que permiten modelar, generar, discretizar o categorizar atributos que intervienen en la creación de modelos.

Cada operador cumple con una función específica y la correcta aplicación de los mismos generará resultados satisfactorios.

A continuación se describen cada uno de los operadores empleados en la generación de cada modelo de minería de datos:

- **Criterios de Interés existentes en el proceso de evaluación docente.**

- **ID3**

Los operadores necesarios fueron (ver figura 1,2):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_1.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y calf\_cualitativa.

**Dis\_calificacion:** operador útil para discretizar la calf\_cualitativa en Criter-Int (Criterios de Interés), Buen-Rend (Buen Rendimiento).

**Dis\_porcentaje:** operador útil para discretizar el porcentaje de calificación obtenida por un docente.

**Numerical to Polynominal:** operador empleado para convertir datos numéricos en polinomiales.

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el desempeño del modelo.

**algoritmo\_ID3:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente observar su desempeño.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.

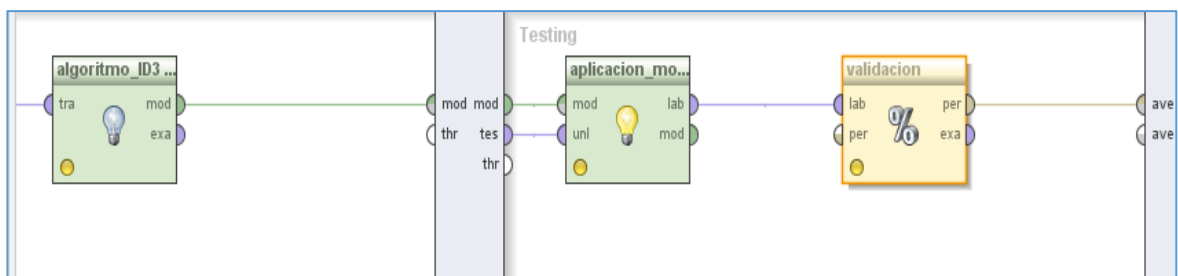


Figura 1: Validación Cruzada ID3.

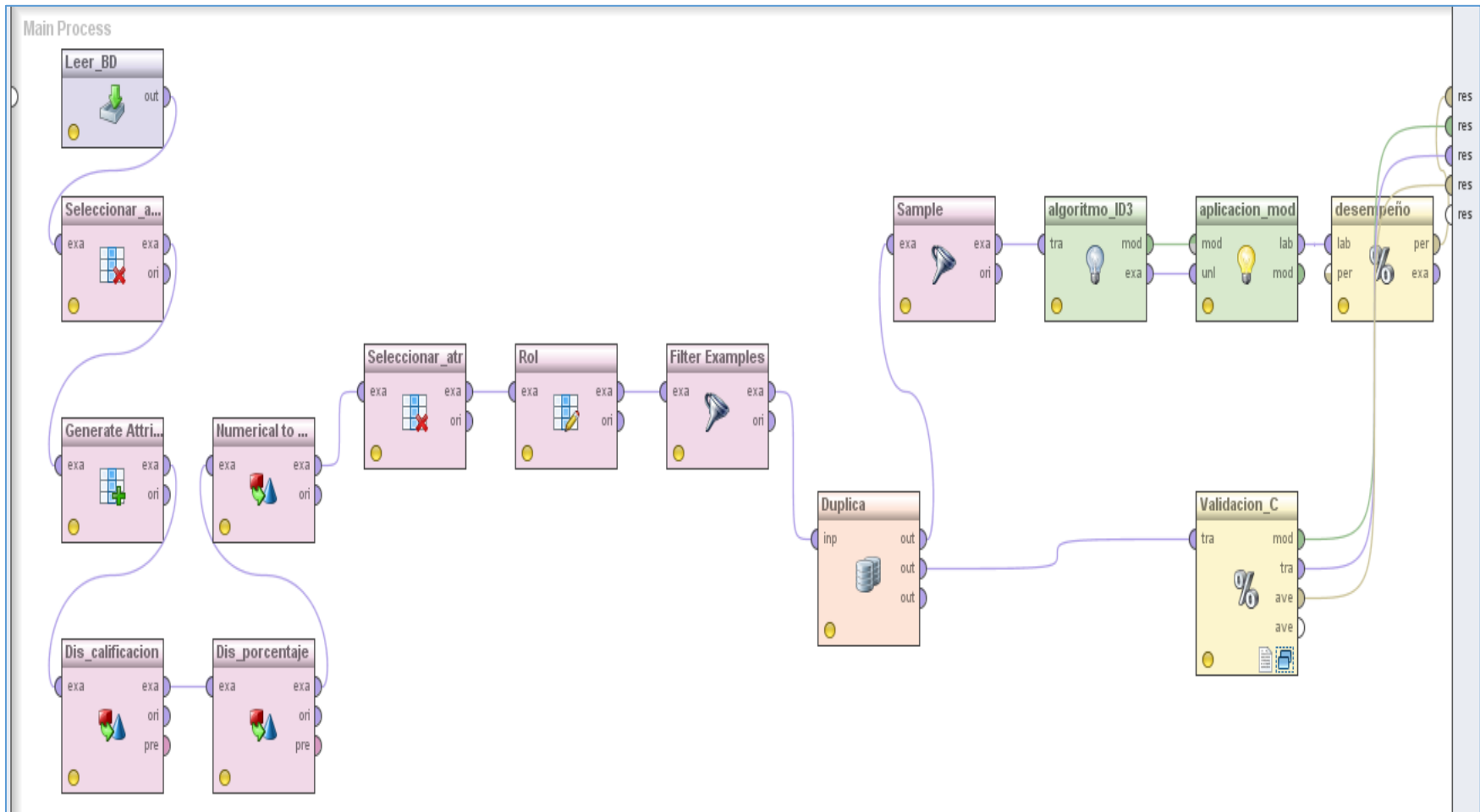


Figura 2: Secuencia de operadores en algoritmo ID3, RapidMiner.

## ➤ CHAID

Los operadores necesarios fueron (ver figura 3,4):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es `app_estructura_1`.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y `calf_cualitativa`.

**Dis\_calificacion:** operador útil para discretizar la `calf_cualitativa` en Criter-Int (Criterios de Interés), Buen-Rend (Buen Rendimiento).

**Dis\_porcentaje:** operador útil para discretizar el porcentaje de calificación obtenida por un docente.

**Numerical to Polynomial:** operador empleado para convertir datos numéricos en polinomiales.

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el desempeño del modelo.

**CHAID:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente observar su desempeño.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.

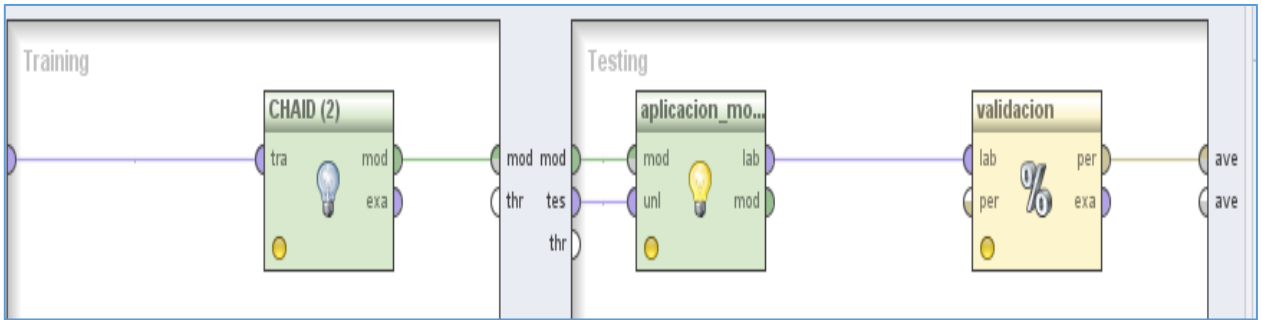


Figura 3: Validación Cruzada CHAID.

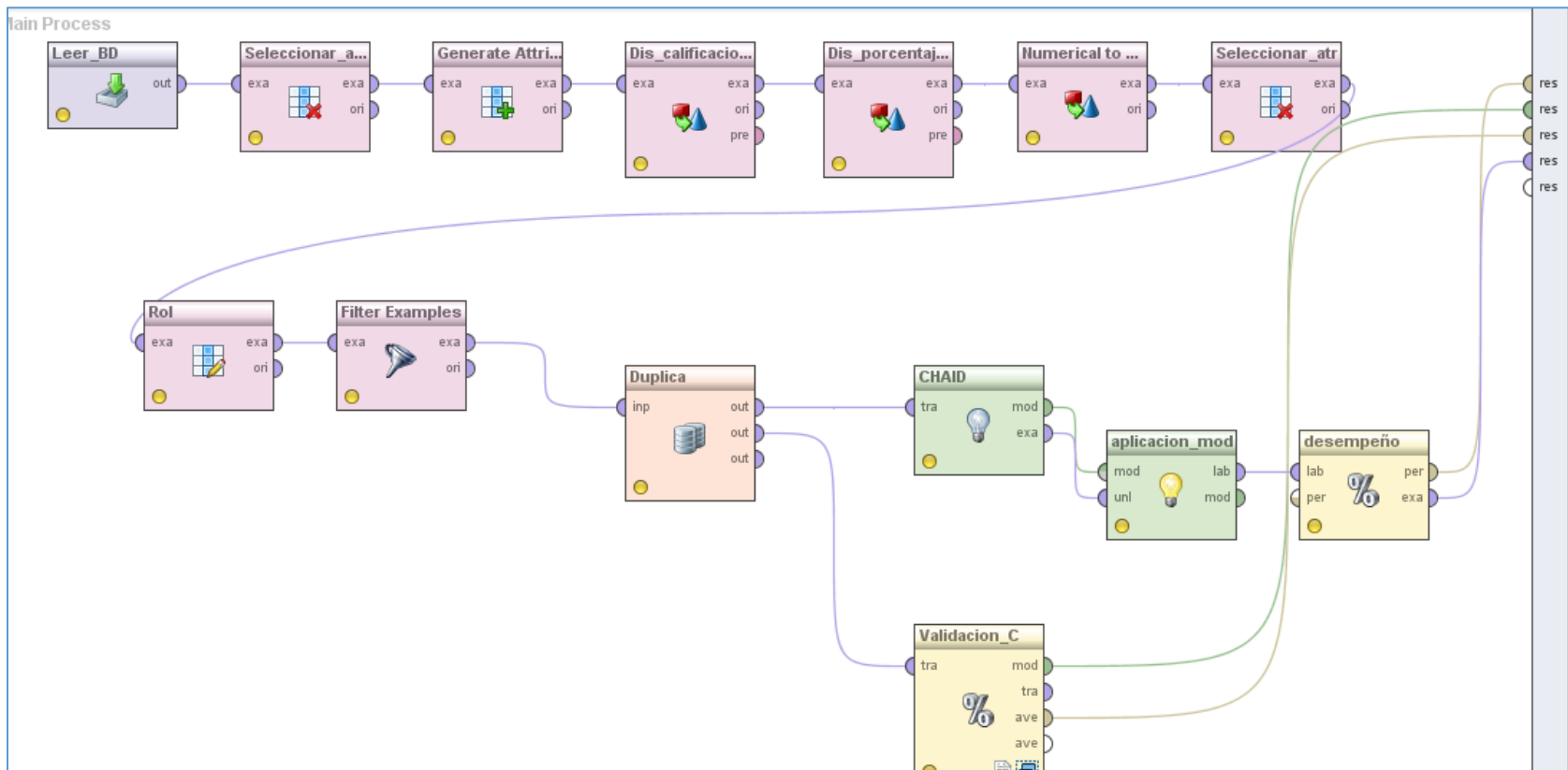


Figura 4: Secuencia de operadores en algoritmo CHAID, RapidMiner.

#### ➤ C4.5

Los operadores necesarios fueron (Ver Figura 5,6):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_1.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y calf\_cualitativa.

**Dis\_calificacion:** operador útil para discretizar la calf\_cualitativa en Criter-Int (Criterios de Interés), Buen-Rend (Buen Rendimiento).

**Dis\_porcentaje:** operador útil para discretizar el porcentaje de calificación obtenida por un docente.

**Numerical to Polynominal:** operador empleado para convertir datos numéricos en polinomiales.

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el desempeño del modelo.

**algoritmo\_C4.5:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente observar su desempeño.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.



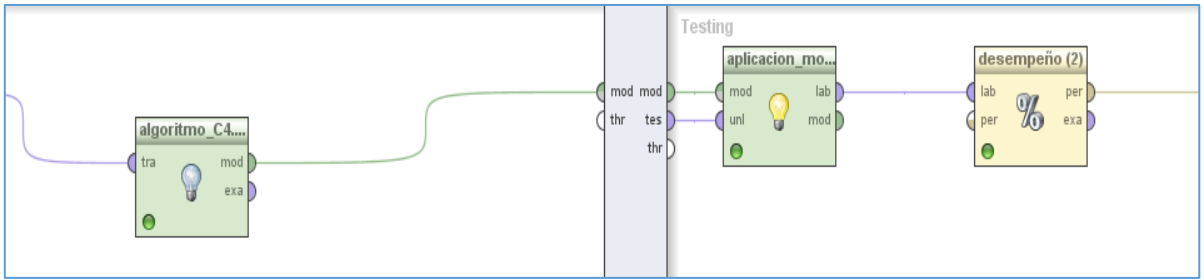


Figura 5: Validación Cruzada

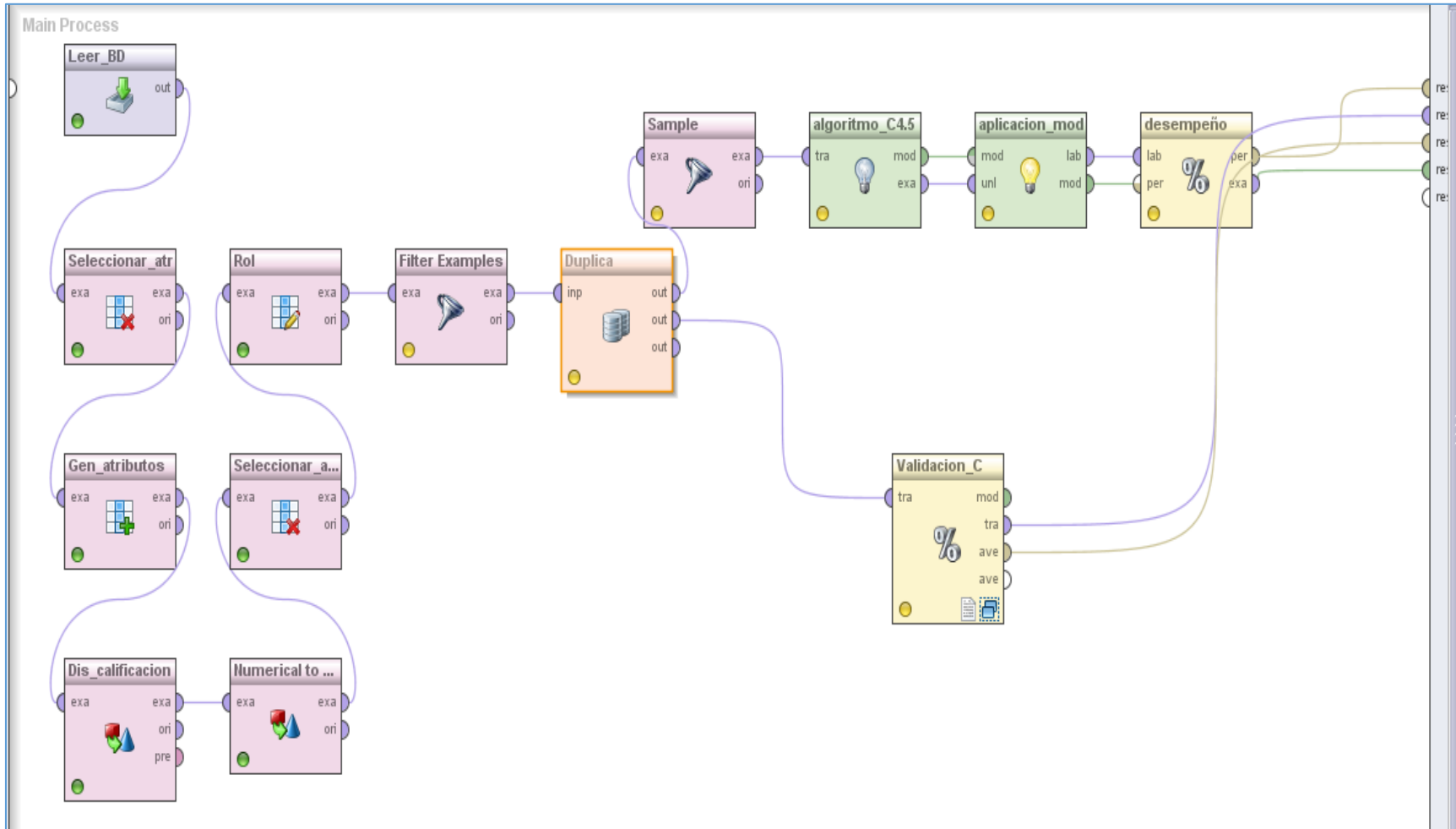


Figura 6: Secuencia de operadores en algoritmo C4.5, RapidMiner.

- **Extracción Opiniones Buen Rendimiento, Criterios Interés**

Para el análisis de las opiniones dentro del proceso de evaluación se procedió a extraer las mismas primero en base a la cedula docente, en donde existan tanto criterios de interés, como opiniones de buen rendimiento docente para luego unir a la tabla la cual contiene todas opiniones. Los operadores utilizados fueron los siguientes (Ver Figura 7, 8):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es `app_estructura_1`.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y `calf_cualitativa`.

**Dis\_calificacion:** operador útil para discretizar la `calf_cualitativa` en Criter-Int (Criterios de Interés), Buen-Rend (Buen Rendimiento).

**Dis\_porcentaje:** operador útil para discretizar el porcentaje de calificación obtenida por un docente.

**Numerical to Polynominal:** operador empleado para convertir datos numéricos en polinomiales

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**Write CSV:** operador útil en la exportación de datos en formato `.csv`

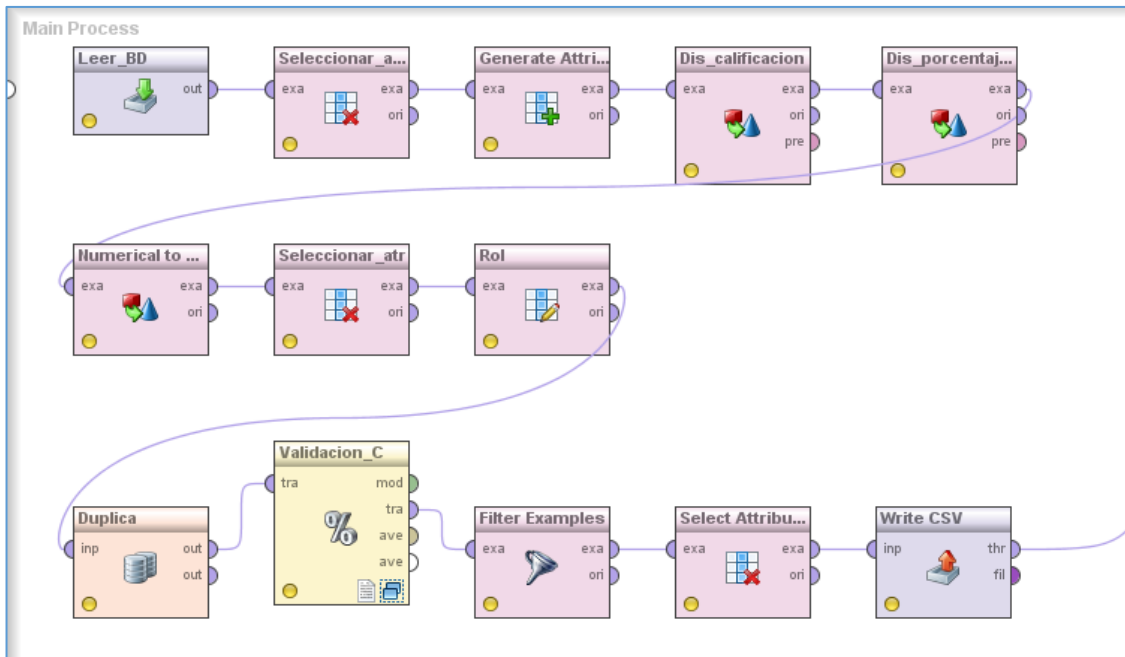


Figura 6: Obtención Docentes: Criterios Interés, Buen Rendimiento

- **Opiniones**

Luego de la obtención de los docentes con criterios de interés o buen rendimiento se procede a unir la tabla de los docentes seleccionados y la tabla que contiene las opiniones dentro del proceso. Los operadores útiles fueron:

**Leer\_BD:** operador empleado para leer los datos de los docentes con la estructura que contiene las opinones app\_est\_c.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Join:** operador útil para unir 2 tablas.

**Write CSV:** operador útil en la exportación de datos en formato .csv

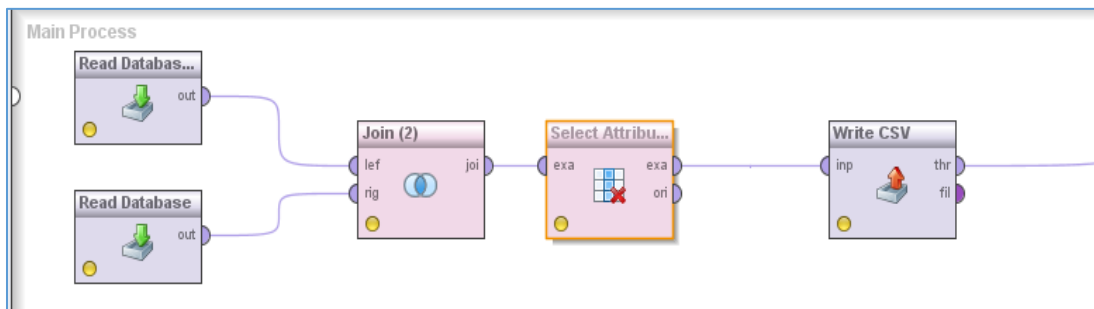


Figura 7: Obtención de Opiniones

- **Evolución de cátedra que ha tenido un docente en los diferentes periodos académicos.**

- **Naive Bayes**

Los operadores necesarios fueron (Ver Figura 8, 9):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_3 y app\_estructura\_2.

**Join:** operador útil para unir 2 tablas.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resul\_cpf, catedra\_cpf entre otros.

**Dis\_calificacion:** operador útil para discretizar la catedra docente en Positiva y Negativa.

**Numerical to Binomial:** operador empleado para convertir datos numéricos en binomiales en este caso true-false.

**RoI:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Naive Bayes:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente observar su desempeño.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.

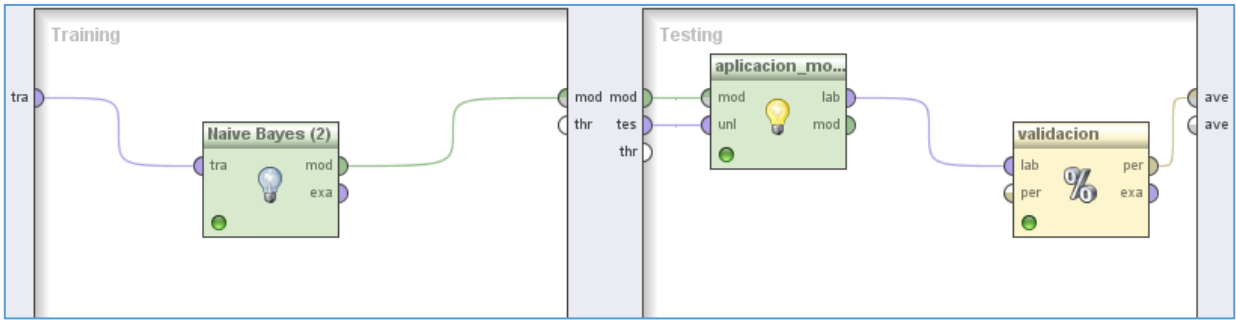


Figura 8: Validación Cruzada.

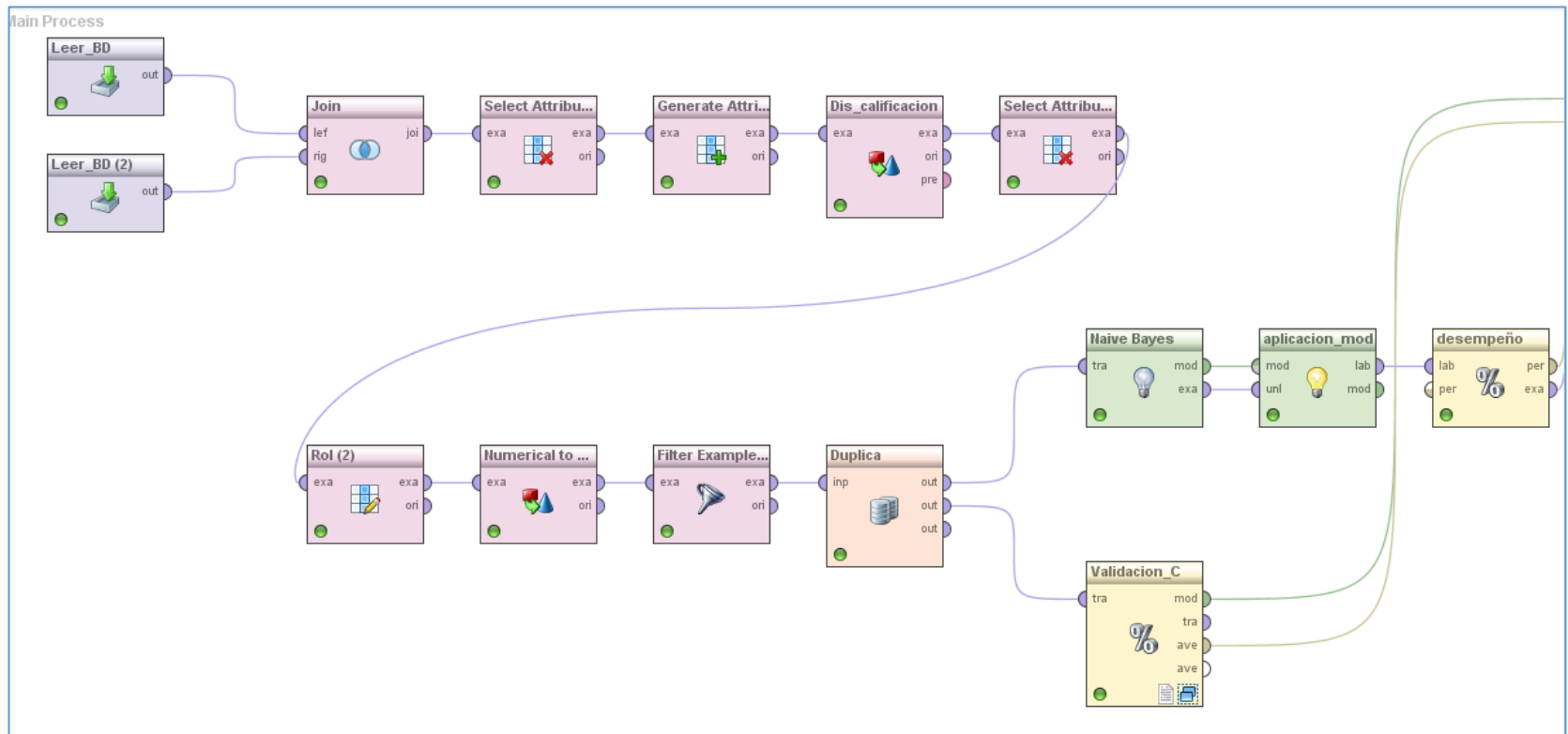


Figura 9: Secuencia de operadores en algoritmo Naive Bayes, RapidMiner

➤ **ID3**

Los operadores necesarios fueron (Ver Figura 10, 11):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_2 y app\_estructura\_3.

**Join:** operador útil para unir 2 tablas.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resul\_cpf, catedra\_cpf entre otros.

**Dis\_calificacion:** operador útil para discretizar la calificación por componente en P (Positiva), N (Negativa).

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Numerical to Binominal:** operador empleado para convertir datos numéricos en binomiales en este caso true-false.

**Filter Examples:** operador útil para filtrar atributos de interés

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**algoritmo\_ID3:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente observar su desempeño.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.



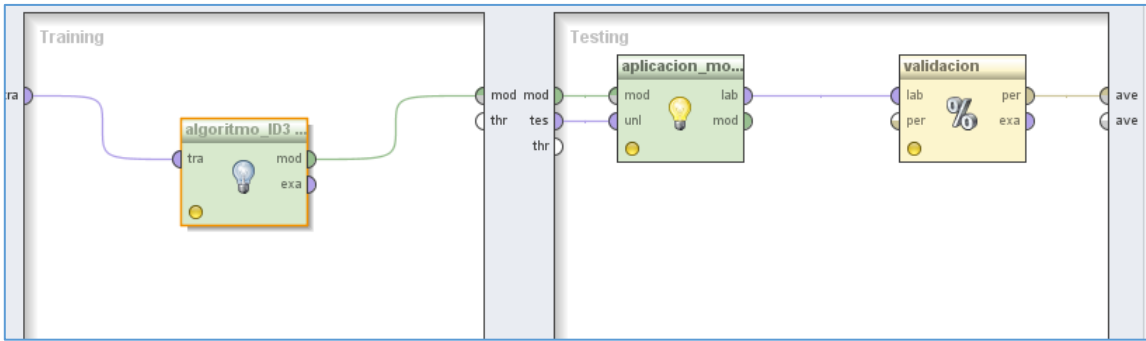


Figura 10: Validación Cruzada.

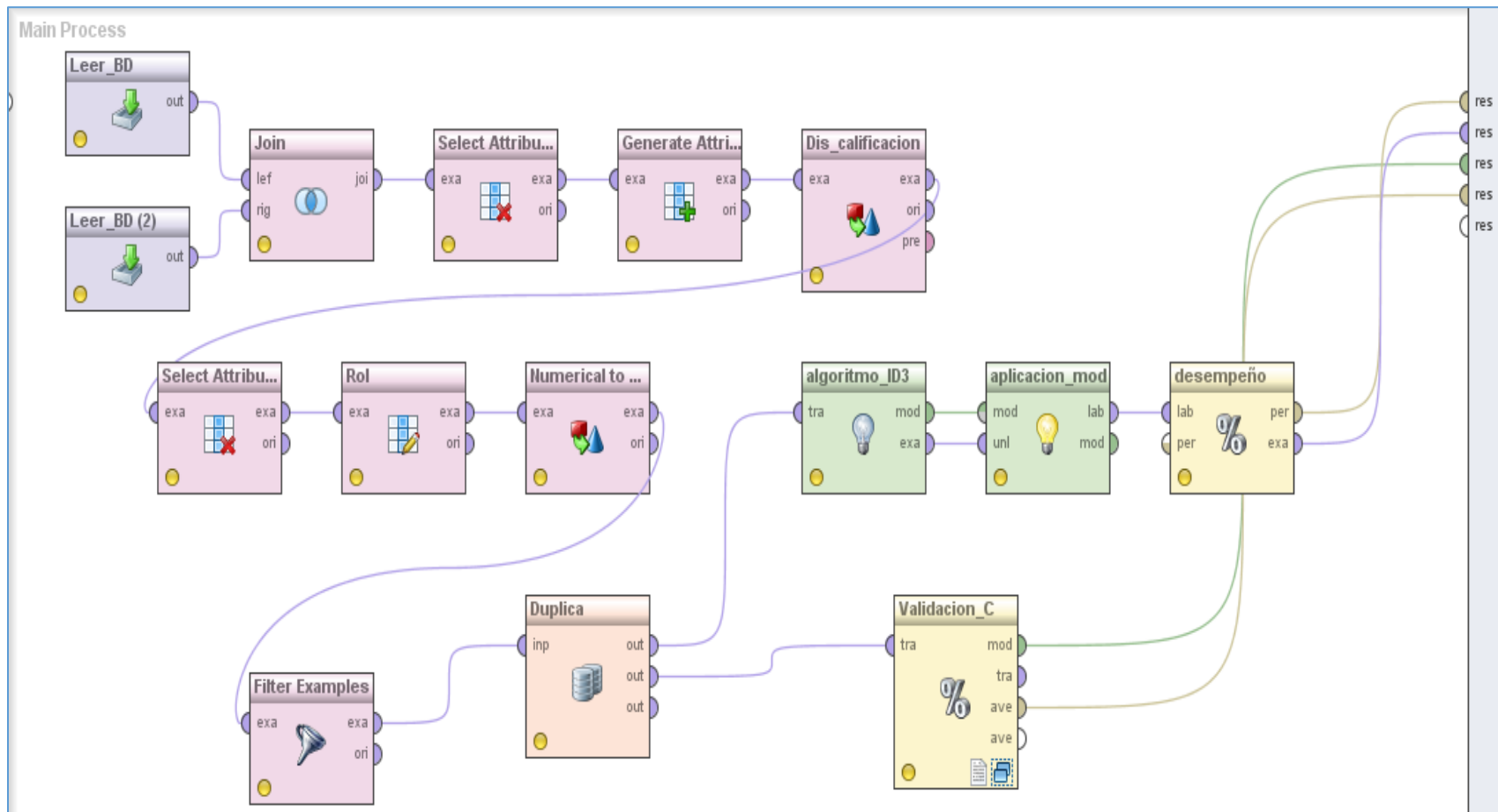


Figura 11: Secuencia de operadores en algoritmo ID3, RapidMiner.

➤ **W-NNge**

Los operadores necesarios fueron (Ver Figura 12, 13):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_3 y app\_estructura\_2.

**Join:** operador útil para unir 2 tablas.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Gen\_atributos:** operador útil en la generación de atributos como: resul\_cpf, catedra\_cpf entre otros.

**Dis\_calificacion:** operador útil para discretizar la calificación por componente en P (Positiva), N (Negativa).

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Numerical to Binominal:** operador empleado para convertir datos numéricos en binomiales en este caso true-false.

**Write Arff:** operador útil en la exportación de datos en formato .arff

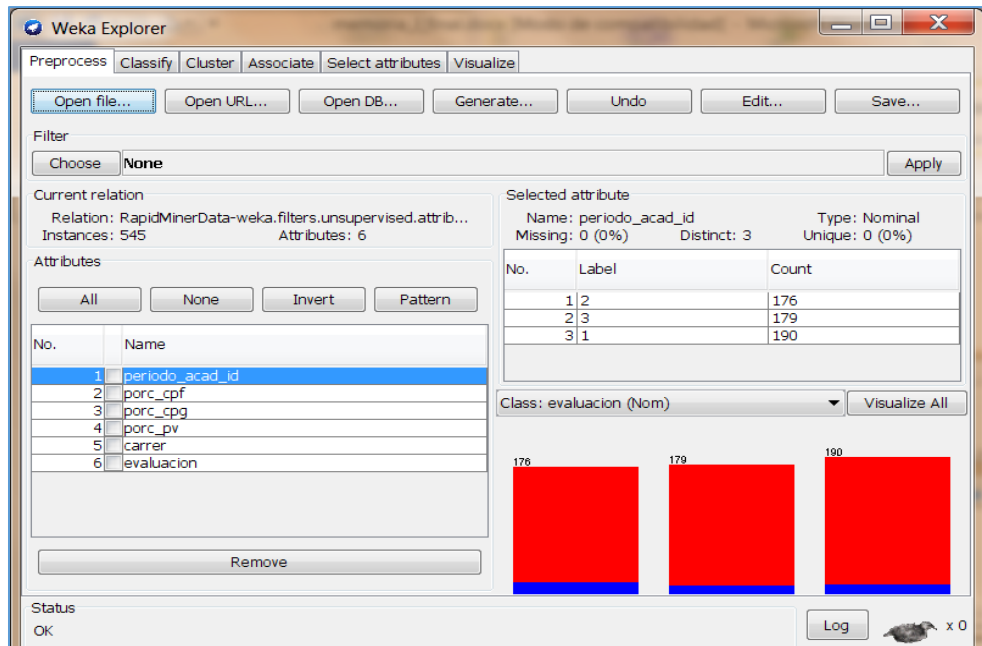


Figura 12: Archivo arff en algoritmo W-NNge, RapidMiner.

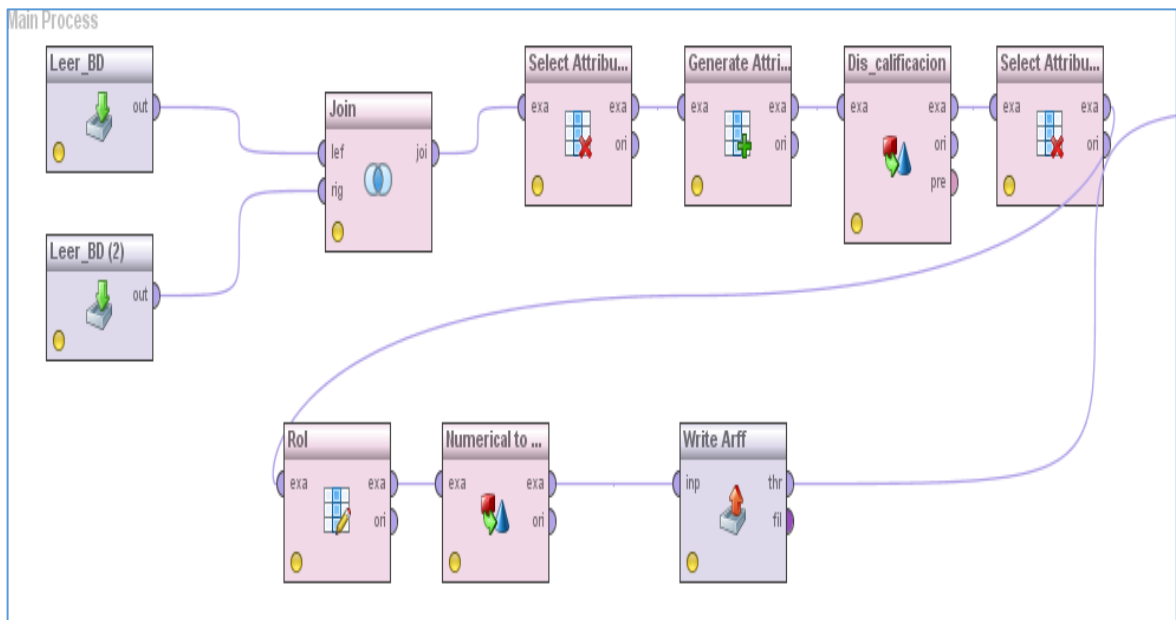


Figura 13: Secuencia de operadores en algoritmo W-NNge, RapidMiner.

- **Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.**

➤ **Regresión Logística:**

Los operadores necesarios fueron (Ver Figura 14, 15):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_1.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y calf\_cualitativa.

**Dis\_calificacion:** operador útil para discretizar la calf\_cualitativa en alta o baja

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el rendimiento del modelo.

**regresion\_log:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente evaluarlo.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.

**Exportar\_Mod:** operador que permite exportar el modelo, para en este caso aplicarlo a otra parte de la base de datos.

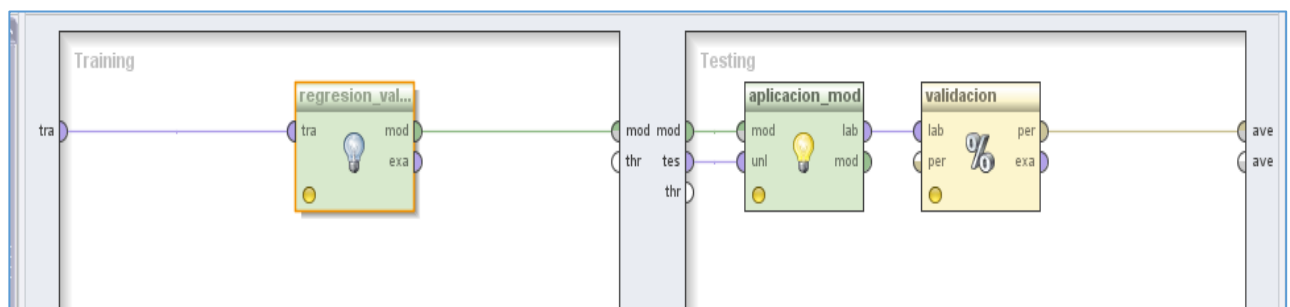


Figura 14: Validación Cruzada.

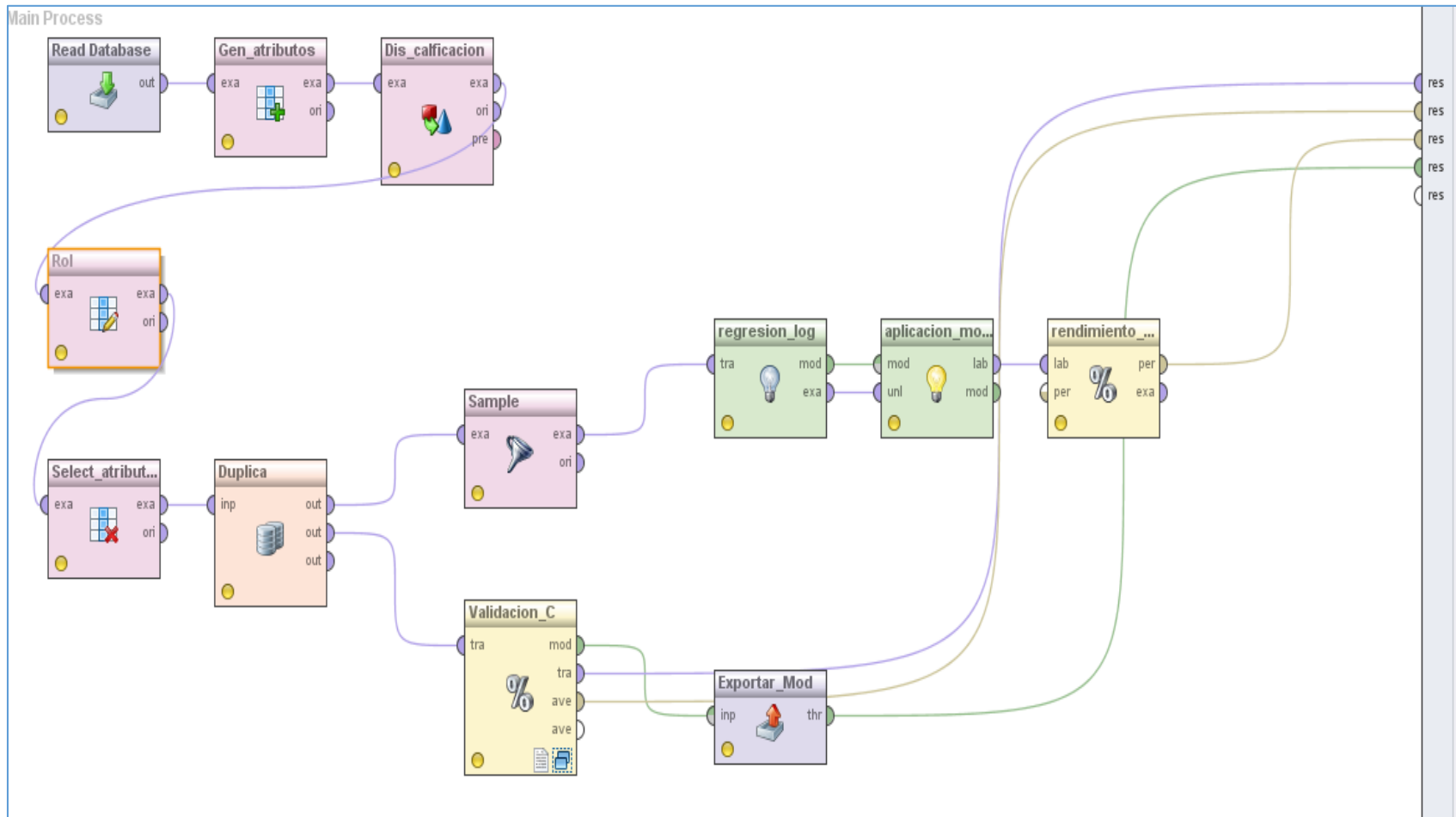


Figura 15: Secuencia de operadores en algoritmo regresión logística, RapidMiner

➤ **Predicción Regresión Logística:**

Los operadores necesarios fueron (Ver Figura 16):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es `app_estructura_1`.

**Leer\_Modelo:** operador que sirve para leer el modelo generado por el algoritmo de regresión logística.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y `calf_cualitativa`.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el rendimiento del modelo.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente obtener la predicción deseada.

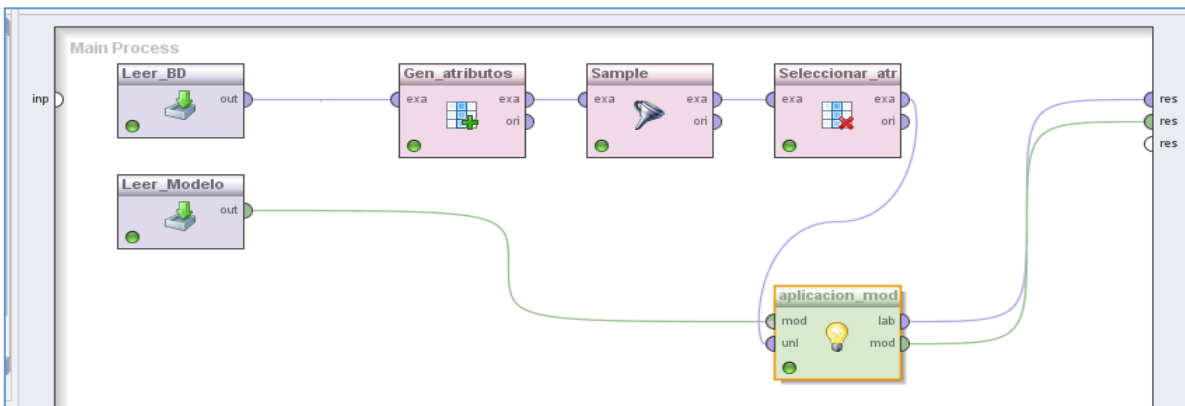


Figura 16: operadores en la predicción de regresión logística

### ➤ ID3

Los operadores necesarios fueron (ver figura 17, 18):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es `app_estructura_1`.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y `calf_cualitativa`.

**Dis\_calificacion:** operador útil para discretizar la `calf_cualitativa` en alta o baja.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Dis\_notas:** operador útil para discretizar las notas de los estudiantes en mala, regular, buena, excelente.

**Numerical to Polynominal:** operador empleado para convertir datos numéricos en polinomiales.

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el desempeño del modelo.

**algoritmo\_ID3:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente evaluarlo.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.



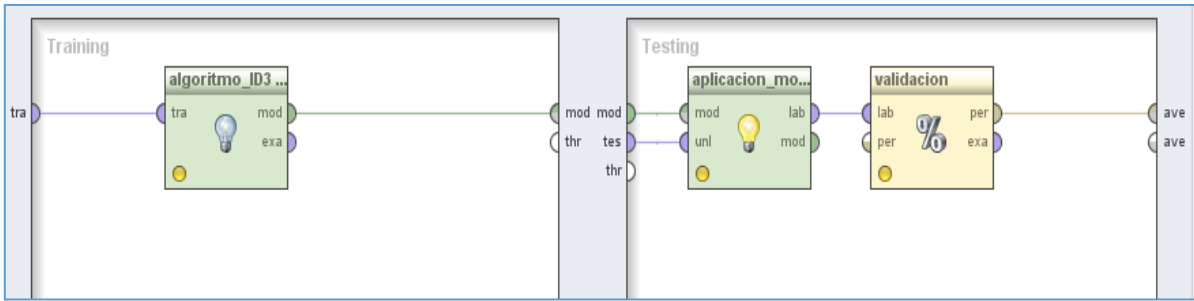


Figura 17: Validación Cruzada.

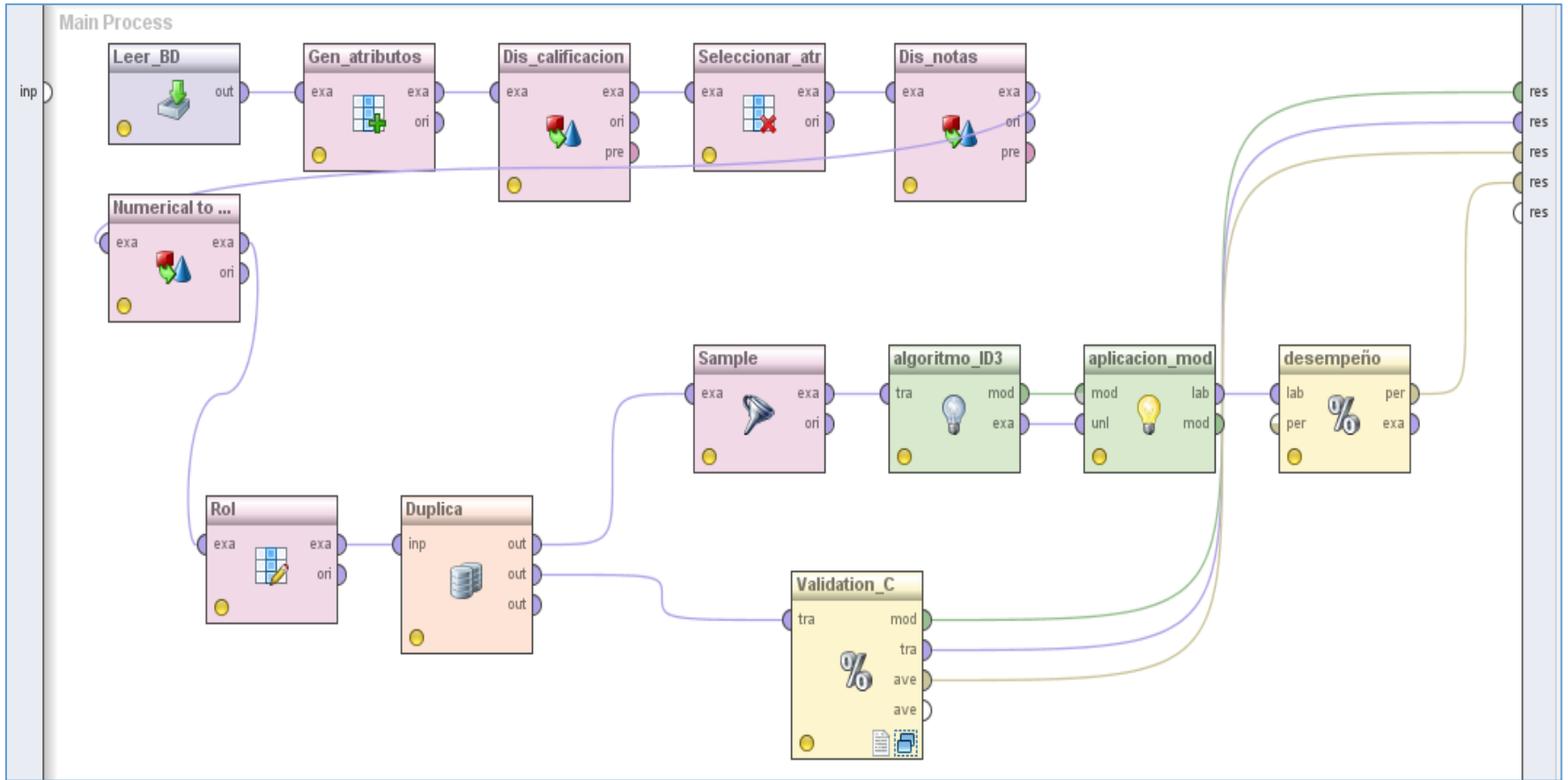


Figura 18: Secuencia de operadores en algoritmo ID3, RapidMiner.

#### ➤ C4.5

Los operadores necesarios fueron (Ver Figura 19,20):

**Leer\_BD:** operador empleado para leer de la base de datos la estructura generada la misma que es app\_estructura\_1.

**Gen\_atributos:** operador útil en la generación de atributos como: resultado, porcentaje y calf\_cualitativa.

**Dis\_calificacion:** operador útil para discretizar la calf\_cualitativa en Baja, Alta.

**Seleccionar\_atr:** este operador ayuda en la selección de atributos útiles en la generación del modelo.

**Dis\_notas:** operador útil para discretizar las notas de los estudiantes en mala, regular, buena, excelente.

**Numerical to Polynomial:** operador empleado para convertir datos numéricos en polinomiales.

**Rol:** operador que asigna roles como id, label y regular para identificar cada registro y atributo a predecir o describir.

**Duplica:** operador empleado para replicar la base de datos con el fin de evaluar el modelo y aplicar una validación cruzada.

**Sample:** operador que toma una muestra de la base de datos en donde se evaluará el desempeño del modelo.

**algoritmo\_C4.5:** operador contiene el algoritmo que se aplica en la generación del modelo, además el mismo operador se usa dentro de la validación cruzada.

**aplicacion\_mod:** operador útil para la aplicación del modelo al conjunto de datos para posteriormente evaluarlo.

**desempeño:** con este operador se evalúa el rendimiento del algoritmo aplicado a los datos seleccionados.

**Validacion\_C:** es un operador que contiene más subprocesos, divide la base de datos subconjuntos, además de realizar la división de dos escenarios una primera parte de entrenamiento y otra de prueba, donde se da un ensayo previo y luego se aplica el modelo a una parte de los datos que no hayan visto el modelo, todo esto con el fin de realizar la validación del modelo generado.

**validacion:** operador que da la validación del modelo generado.

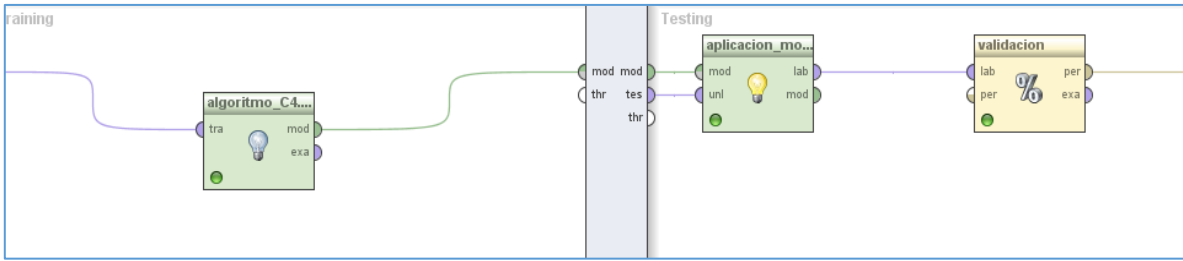


Figura 19: Validación Cruzada.

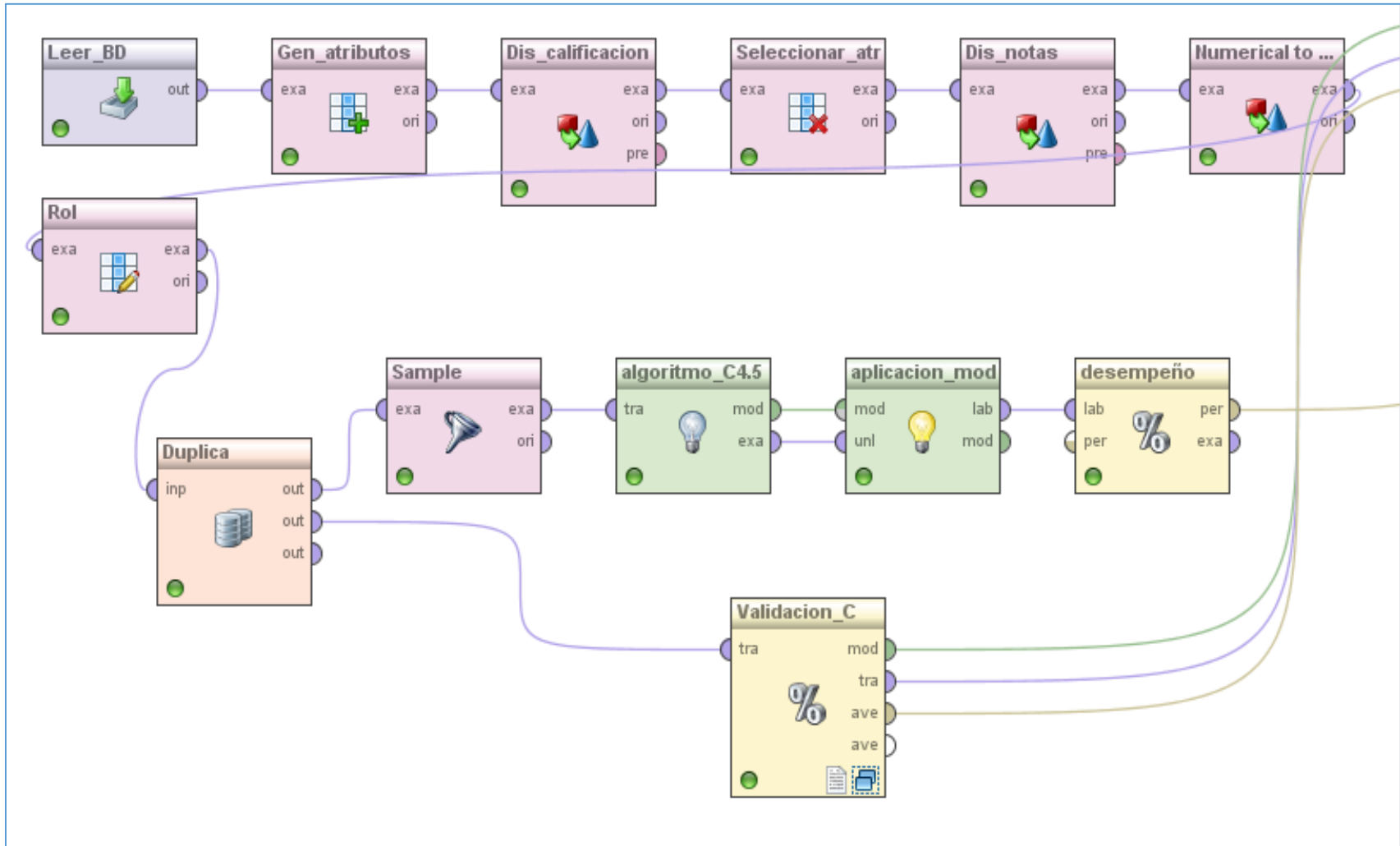


Figura 20: Secuencia de operadores en algoritmo C4.5, RapidMiner.

## ANEXO 8: Análisis General: Escenario Real

- **Análisis: Criterios de Interés existentes en el proceso de evaluación docente**

En este aspecto se puede observar que las carreras del AEIRNNR poseen docentes en su mayoría con un buen rendimiento, existiendo a la vez en todos los periodos de evaluación criterios de interés que pueden reforzar el plan de mejoras dirigido hacia un docente (Ver Figura 1)

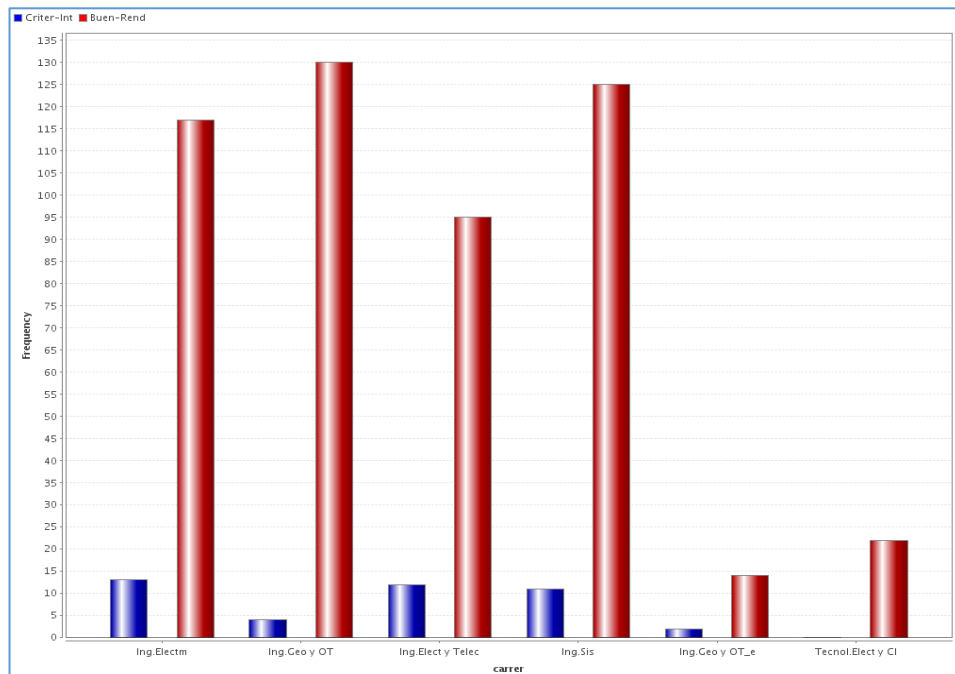


Figura 1: Criterios de Interés, Buen Rendimiento

El gráfico anterior también evidencia que dentro de la carrera de Tecnología en Electricidad y Control Industrial no existen criterios de interés, pero las carreras de Ingeniería en Electromecánica, Ingeniería en Sistemas e Ingeniería en Electrónica y Telecomunicaciones son las carreras con más criterios de interés presenta en los periodos de evaluación con 8.08%, 10%, 11.21% respectivamente, además la carrera con mayor cantidad de docentes con buen rendimiento es la carrera de Geología Ambiental y Ordenamiento Territorial con un 92.01% del total de docente que corresponden a dicha carrera (Ver Figura 2).

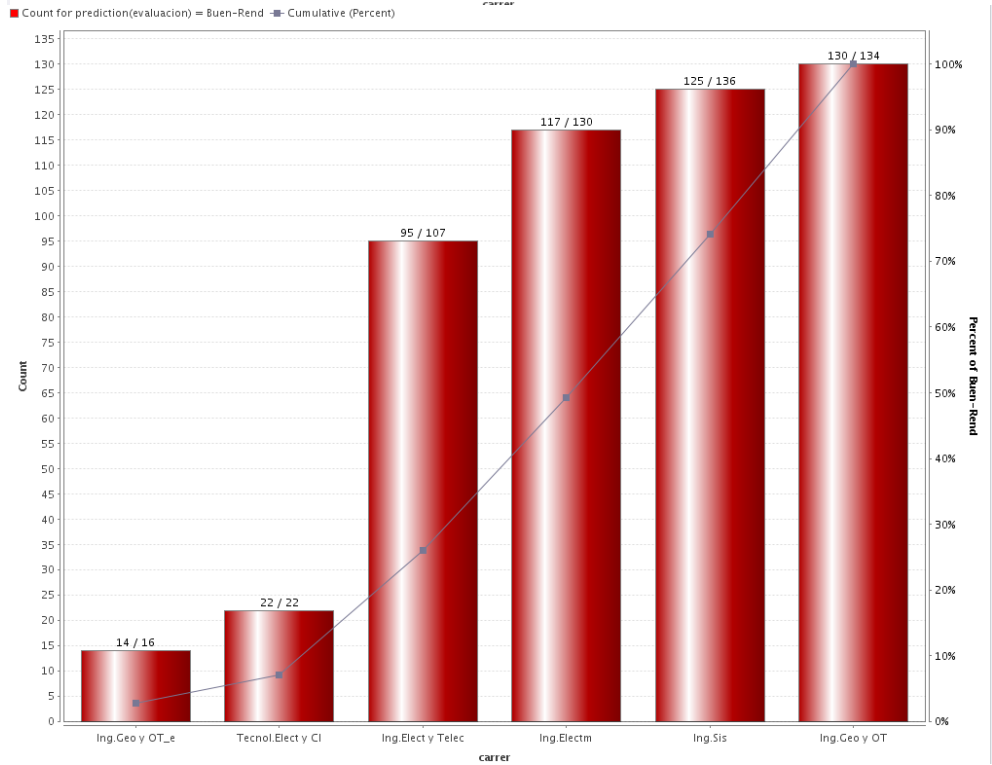
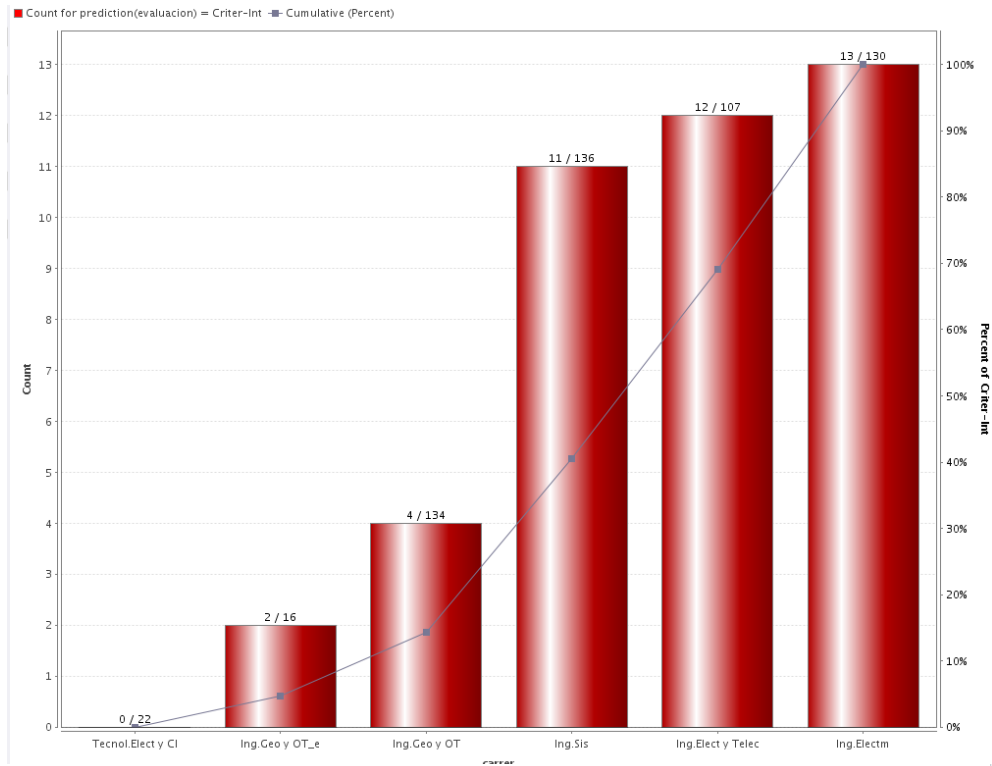


Figura 2: Cantidad Docentes con Buen rendimiento y Criterios de Interés

Las carreras con menos docentes con criterios de interés en el proceso de evaluación, son las carreras de Ing. en Geología Ambiental y Ordenamiento Territorial, presencial y especial, seguido de la carrera de Tecnología en Electricidad y Control Industrial.

La figura anterior refuerza lo evaluado e interpretado en la Sección Resultados: Validación del Modelo y Análisis de Resultados, efectivamente existen opiniones vertidas por los estudiantes dentro de un periodo de evaluación, que al ser tomadas en cuenta pueden ser de apoyo a la toma de decisiones en cuanto a la elaboración de un plan de mejoras que se ajuste a los requerimientos que presenta el sistema, tanto para un buen rendimiento, como en la existencia de criterios de interés.

A continuación se presenta un fragmento de las opiniones más relevantes vertidas por los estudiantes dentro del proceso tanto para la existencia de criterios de interés como para el buen rendimiento obtenido por los docentes:

Las opiniones que resaltan más cuando existen criterios de interés son las siguientes (Ver Tabla CX):

**TABLA CX: CRITERIOS DE INTERÉS MAYOR NÚMERO DE OCURRENCIAS**

<b>Palabra</b>	<b>Ocurrencia</b>	<b>Contexto</b>
enseñanza	143	<ul style="list-style-type: none"> <li>• Que el docente mejore en algunos aspectos relacionados con la teoría y enseñanza.</li> <li>• Se debe explicar mejor los temas de la materia, que se escuche cuando un alumno comunica que no entiende las explicaciones, mejorar la enseñanza.</li> <li>• La enseñanza que sigue el docente no va de acuerdo a los intereses y objetivos de los estudiantes.</li> <li>• Enseñanza. Que cambie la forma de enseñar falta planificación</li> </ul>
mejorar	140	<ul style="list-style-type: none"> <li>• Mejorar su pedagogía, enseñanza, capacidad profesional.</li> <li>• Mejorar la comunicación docente-alumno.</li> </ul>
falta	103	<ul style="list-style-type: none"> <li>• Falta de preparación de las clases y falta del cumplimiento del sílabo.</li> <li>• Falta de comprensión docente-alumno, que el proceso de enseñanza sea más estricto para que los trabajos se realicen con más claridad.</li> <li>• Falta de planificación.</li> <li>• Falta de explicación</li> <li>• Falta de Interacción docente-alumno.</li> <li>• Falta de dominio en los contenidos.</li> <li>• Falta metodología de enseñanza</li> </ul>
conocimientos	87	<ul style="list-style-type: none"> <li>• Que se implemente una nueva metodología para cimentar de una mejor manera los conocimientos.</li> <li>• Explicar los contenidos elegidos, mayor dominio de los conocimientos.</li> <li>• Prepararse mejor para que se pueda explicar correctamente los conocimientos.</li> <li>• Los conocimientos adquiridos, son pocos en relación a los que se esperaba,</li> </ul>



		no se trata temas a profundidad.
mayor	62	<ul style="list-style-type: none"> <li>• Mayor compromiso a la hora de dar clases, los trabajos personales deben ser tratados fuera de clases no dentro de ella.</li> <li>• Mayor información sobre la bibliografía de los contenidos estudiados.</li> <li>• Mayor cumplimiento con los horarios de clases.</li> <li>• Mayor preparación del docente antes de dictar clases, y mayor explicación.</li> </ul>
ser	111	<ul style="list-style-type: none"> <li>• Ser más tolerante en las acreditaciones, un estudiante está en formación y no se encuentra al mismo nivel del profesional.</li> <li>• Ser más comprensivo, dar asesoría extra-clase.</li> <li>• Ser más responsable con las actividades académicas, buscar la forma adecuada de impartir los conocimientos, buscar nuevos métodos de enseñanza.</li> <li>• Ser más accesible.</li> <li>• Ser más ordenado en cuanto al desenvolvimiento de la unidad.</li> </ul>
dar	97	<ul style="list-style-type: none"> <li>• Dar clases en donde se informe bien sobre los temas a tratar.</li> <li>• Dar clases impartidas por el docente, no por los alumnos.</li> <li>• Dar una mejor explicación de la materia.</li> </ul>
mejore	71	<ul style="list-style-type: none"> <li>• Mejore su metodología de enseñanza.</li> <li>• Mejore su ética profesional.</li> <li>• Mejore su capacidad profesional y se actualice con las tecnologías de información.</li> <li>• Mejore la puntualidad.</li> </ul>
regular	48	<ul style="list-style-type: none"> <li>• Regular: crear estrategias de enseñanza-aprendizaje.</li> <li>• Desempeño docente regular.</li> </ul>
prepare	54	<ul style="list-style-type: none"> <li>• Que se prepare, para que el docente tenga mayor dominio de la materia.</li> <li>• Que se prepare ya que no cuenta con un método de enseñanza.</li> <li>• Preparar mejor las clases, ya que los temas no son tratados con profundidad.</li> </ul>
tiempo		<ul style="list-style-type: none"> <li>• Tiempo para que el estudiante practique lo enseñado.</li> <li>• Tiempo para las demás materias, desarrollar los trabajos extra-clase.</li> <li>• Tiempo por parte del docente al momento de impartir clases.</li> </ul>
debería	16	<ul style="list-style-type: none"> <li>• El docente debería, aceptar sugerencias de estuantes, dar más explicación, utilizar nuevos métodos de enseñanza, actualizarse.</li> </ul>
aprendizaje	61	<ul style="list-style-type: none"> <li>• Que el docente elabore material didáctico para fortalecer el aprendizaje además de crear estrategias enseñanza-aprendizaje.</li> <li>• Más comunicación profesor-alumno para mejorar los temas tratados en clase, lo cual daría un mejor aprendizaje.</li> <li>• Es necesario que los docentes estén en constante aprendizaje.</li> </ul>
estudiante	97	<ul style="list-style-type: none"> <li>• Mayor comunicación con el estudiante.</li> <li>• Buscar nuevas formas de enseñanza para llegar al estudiante.</li> </ul>
tener		<ul style="list-style-type: none"> <li>• Tener más comunicación con los estudiantes.</li> <li>• Tener más humildad y amabilidad con respecto al trato con los estudiantes.</li> <li>• Tener un trato igual para con todos los alumnos.</li> </ul>

Las opiniones que resaltan más cuando al buen rendimiento docente son las siguientes (Ver Tabla CXI):

**TABLA CXI: CRITERIOS BUEN RENDIMIENTO- MAYOR NÚMERO DE OCURRENCIAS**

Palabra	Ocurrencia	Contexto
bien	385	<ul style="list-style-type: none"> <li>• Bien: promueve la investigación en los alumnos.</li> <li>• Bien: siga cumpliendo de la misma forma su trabajo.</li> </ul>

		<ul style="list-style-type: none"> <li>• Bien: muy buenos logros, enseña a ser una mejor persona.</li> <li>• Bien: escucha al estudiante.</li> <li>• Bien: ha sido muy ético, se comprendió la totalidad de la materia.</li> </ul>
buenos	117	<ul style="list-style-type: none"> <li>• Presenta buenos valores éticos y morales.</li> <li>• Buenos valores, buen profesor éticamente un gran profesional.</li> <li>• Buenos conocimientos adquiridos a través del docente.</li> <li>• Buenos logros, gracias al docente.</li> <li>• Buenos logros académicos con la metodología aplicada</li> </ul>
buena	176	<ul style="list-style-type: none"> <li>• Buena relación con los estudiantes.</li> <li>• Buena metodología de enseñanza, las clases son dinámicas, existe conocimientos.</li> <li>• Buena ética profesional.</li> <li>• Buena persona, califica adecuadamente.</li> <li>• Buena persona tanto personal como profesional.</li> </ul>
excelente	194	<ul style="list-style-type: none"> <li>• Excelente docente, persona, profesor, profesional se preocupa por los estudiantes, responsable en lo que realiza.</li> <li>• Excelente comportamiento es un gran aporte para la carrera.</li> </ul>
conocimientos	215	<ul style="list-style-type: none"> <li>• Satisfacción en todos los conocimientos adquiridos.</li> <li>• Buenos y nuevos conocimientos.</li> </ul>
comportamiento	112	<ul style="list-style-type: none"> <li>• Comportamiento ético del docente dentro de clases.</li> <li>• El docente tiene un comportamiento muy bueno.</li> </ul>
persona	78	<ul style="list-style-type: none"> <li>• El docente es una persona muy respetuosa, responsable, justa, agradable, que transmite su buena ética y moral a los estudiantes.</li> </ul>
valores	85	<ul style="list-style-type: none"> <li>• El docente tiene una buena práctica de valores.</li> <li>• El docente enseña buenos valores.</li> </ul>
enseñanza	110	<ul style="list-style-type: none"> <li>• El docente sabe enseñar, es excelente ya que las clases son progresivas, por lo tanto posee buena destreza y buena metodología de enseñanza.</li> </ul>

- **Análisis: Resultados, Evolución de cátedra que han tenido los docentes en los diferentes periodos académicos.**

La evolución de cátedra que han tenido los docentes en todas las carreras del AEIRNNR presenta una estadística muy grande, las mismas que evidencian el nivel de preparación y profesionalidad existente en cada una de ellas (ver Figura 3, 4).

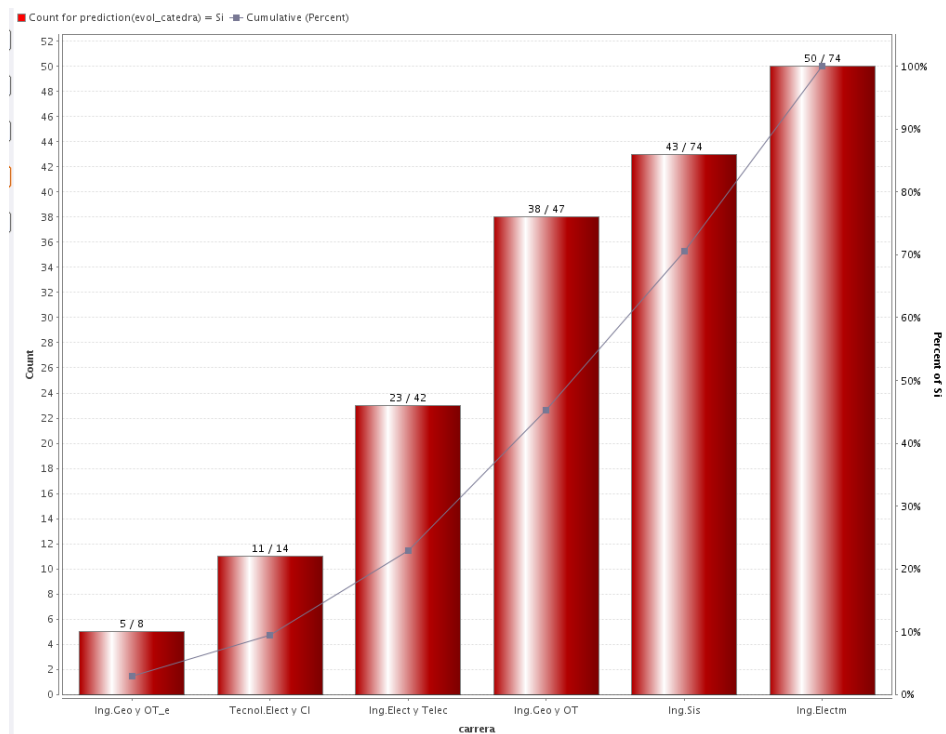


Figura 3: Docentes que presentan una evolución de cátedra

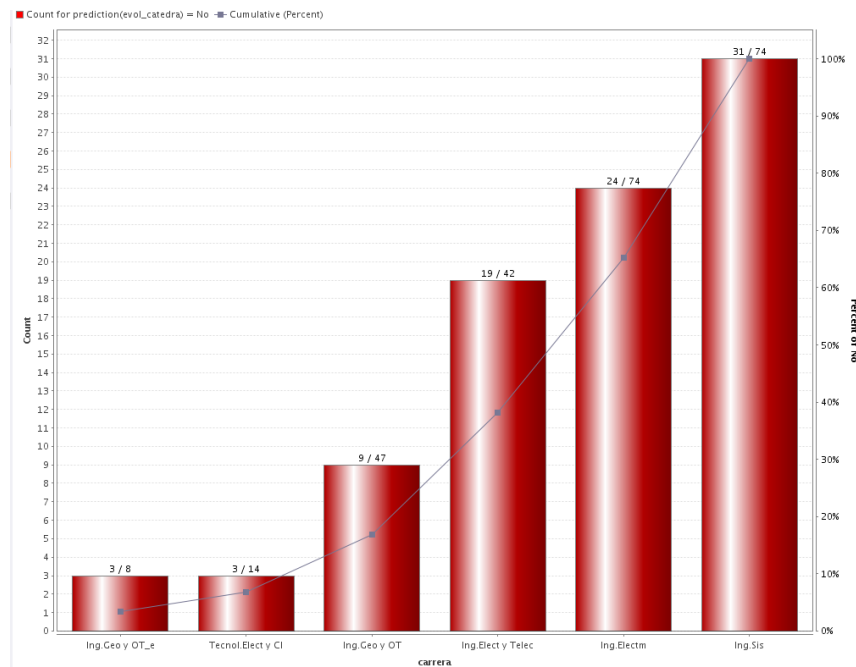


Figura 4: Docentes que No presentan una evolución de cátedra

En las figuras se muestra la cantidad de docentes que presentan una evolución de cátedra por carrera, como aquellos que no la poseen. Se puede observar que todas las carreras tienen un buen número de docentes con evolución de cátedra, siendo un aspecto fundamental en la calidad de la educación superior. En la obtención de una buena calidad en la educación es necesario contar con docentes que a medida pase el tiempo vayan evolucionando y adquiriendo más experiencia tanto profesional como personal, esto se cumple en su mayoría dentro del AEIRNNR.

Pero cabe resaltar que las carreras con más docentes con evolución de cátedra son las carreras de Geología Ambiental y Ordenamiento Territorial, seguido de la carrera de Tecnología en Electricidad y Control Industrial con un 80.85% y 78.57% respectivamente.

Las carreras con menos docentes sin dicha evolución son las carreras de Ingeniería en Electrónica y Telecomunicaciones e Ingeniería en Sistemas, con un 45.23% y 41.89% del total de docentes de dichas carreras.

Este análisis refuerza también refuerza el factor crítico estudiando anteriormente que son los criterios de interés ya que las carreras que presentan menos evolución de cátedra anteriormente también presentaron más criterios de interés.

Lo descrito se puede evidenciar en la Sección Resultados: Validación del Modelo y Análisis de Resultados.

- **Análisis: Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.**

La incidencia de las notas de los estudiantes también se analizaron en la presente investigación, lo cual constituye una fuente importante en la toma de decisiones académicas, la siguiente tabla presenta de manera resumida lo obtenido por los modelos generados:

### Módulos Impares

<b>Calificación Alta</b>			
<b>Nota</b>	<b>Carrera</b>	<b>Módulos</b>	<b>Porcentaje</b>
Excelente	Electrónica y Telecomunicaciones	5, 9	94.80%, 97.43%
	Electromecánica	3	100%
	Geología Ambiental y Ordenamiento T.	7	100%
	Sistemas	5	95.31%
Buena	Electrónica y Telecomunicaciones	5	92.64%
	Electromecánica	7	100%
	Geología Ambiental y Ordenamiento T.	3	99.33%
	Sistemas	5	90.06%
<b>Calificación Baja</b>			
Regular	Electrónica y Telecomunicaciones	7	24.52%
	Electromecánica	3	10.25%

	Geología Ambiental y Ordenamiento T.	7	13.79%
	Sistemas	9	34.42%
Mala	Electrónica y Telecomunicaciones	3	93.10%
	Electromecánica	3	91.30%
	Geología Ambiental y Ordenamiento T.	3	75%
	Sistemas	3	89.36%

### Módulos Pares

<b>Calificación Alta</b>			
<b>Nota</b>	<b>Carrera</b>	<b>Módulos</b>	<b>Porcentaje</b>
Excelente	Electrónica y Telecomunicaciones	10, 6, 8	100%, 98.11%, 100%
	Electromecánica	6, 4	96.96%, 97.93%
	Geología Ambiental y Ordenamiento T.	2, 8	100%, 94.73%
	Sistemas	4, 10	98.07%, 96.77%
Buena	Electrónica y Telecomunicaciones	10,1	100%, 92.40%
	Electromecánica	10, 6	91.81%, 100%
	Geología Ambiental y Ordenamiento T.	6, 8	96.93%, 98%
	Sistemas	10, 8	96.22%, 95.74%
<b>Calificación Baja</b>			
Regular	Electrónica y Telecomunicaciones	2, 1	22.5%, 31.81%
	Electromecánica	4, 8	26.86%, 40%
	Geología Ambiental y Ordenamiento T.	4, 1	31.25%, 13.95%
	Sistemas	6, 6	14.66%, 30%
Mala	Electrónica y Telecomunicaciones	2, 8	86.95%, 95.23%

	Electromecánica	4	78.04%, 95.65%
	Geología Ambiental y Ordenamiento T.	4, 8	81.25%, 84.21%
	Sistemas	2, 1	95.23%, 88.49%

Las tablas muestran la influencia de las notas de los estudiantes en las calificaciones docentes, tanto para módulos pares como para módulos impares y el porcentaje de incidencia que estas tienen en cada carrera.

**Se puede rescatar 3 aspectos:**

- En los módulos impares los estudiantes de las carreras de Electromecánica y Geología Ambiental y Ordenamiento T, que tienen una nota entre 7.6 a 10 califican Alto a los docentes dentro de la evaluación. En cambio los estudiantes que pertenecen a las carreras de Electrónica y Telecomunicaciones, y Sistemas, que tienen una nota entre 0 a 7.5 califican Bajo a los docentes dentro de la evaluación.
- En los módulos pares los estudiantes de las carreras de Electrónica y Telecomunicaciones y Geología Ambiental y Ordenamiento T, que tienen una nota entre 7.6 a 10 califican Alto a los docentes dentro de la evaluación. En cambio los estudiantes que pertenecen a las carreras de Electrónica y Telecomunicaciones, y Sistemas, que tienen una nota entre 0 a 7.5 califican Bajo a los docentes dentro de la evaluación.
- Los módulos donde los estudiantes califican más alto de dependiendo de sus notas son los módulos 5, 6,8 y 10. Los módulos donde los estudiantes califican más bajo dependiendo de sus notas son los módulos 3 y 4 en todas las carreras, en la siguiente figura se puede ver de forma general la influencia de notas en calificaciones docentes (Ver Figura 4).

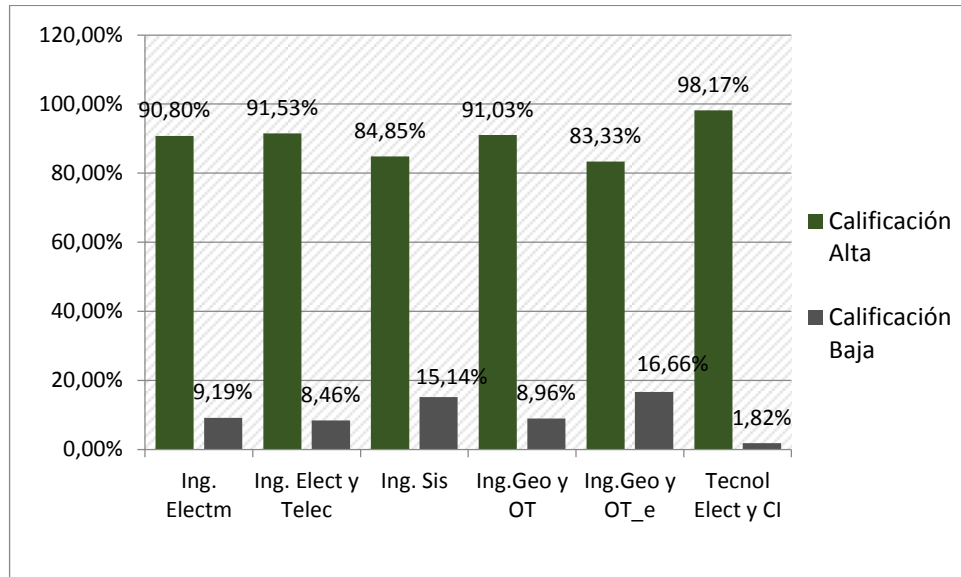


Figura 4: Influencia de Notas en Calificaciones AEIRNNR

En el gráfico se puede observar que todas las carreras del AEIRNNR, tienen docentes en cada asignatura calificados con un margen alto, sin embargo y como se mostró anteriormente en el análisis de resultados, claramente en la evaluación de desempeño de un docente juega un papel muy importante las notas de los estudiantes en los resultados finales de la misma (Ver Figura 5, 6).

Se puede interpretar además que la mayor parte de docentes que posee calificaciones altas es la carrera de Tecnología en Electricidad y Control Industrial con un 98.17%, seguido de Ingeniería en Electrónica y Telecomunicaciones con 91.53%.

Las carreras de Ingeniería en Geología Ambiental y Ordenamiento Territorial y la Carrera de Ingeniería en Sistemas poseen más calificaciones bajas con un 15.14%, 16.66% respectivamente.



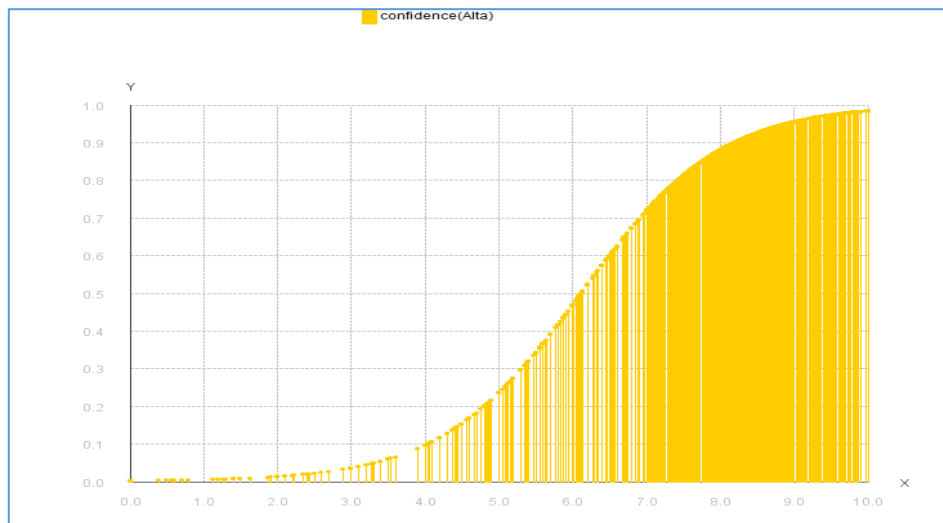


Figura 5: Influencia de Notas Altas en Calificaciones Docentes

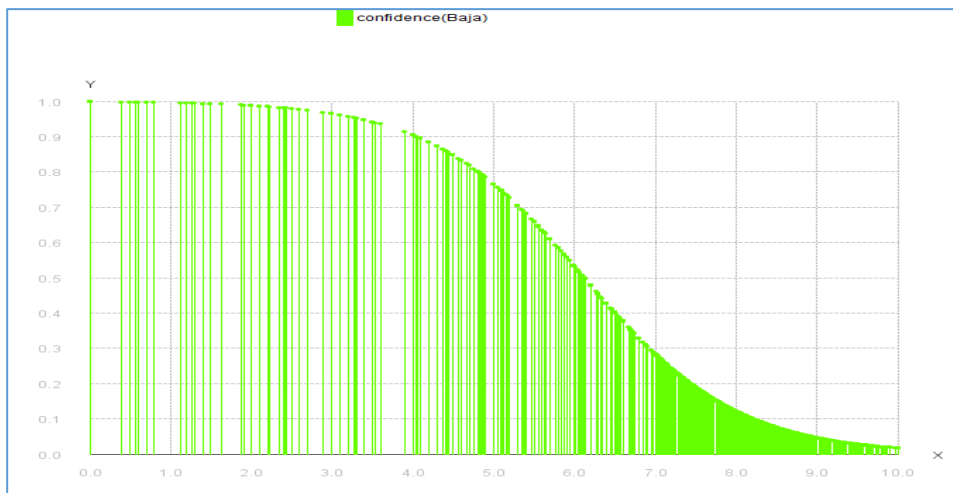


Figura 6: Influencia de Notas Bajas en Calificaciones Docentes

En el gráfico se puede reforzar el análisis realizado con los diferentes algoritmos y modelos generados, es decir a mayor nota de un estudiante, más es la calificación de un docente en la evaluación de desempeño. Esto ocurre en todas las carreras dando un patrón de comportamiento a mayor nota estudiante, entonces alta calificación docente y a menor nota estudiante, entonces baja calificación docente. (Ver Sección Resultados: Validación del Modelo y Análisis de Resultados.)

## ANEXO 9: Nuevos Datos Evaluación De Desempeño Docente

Loja, 31 de Julio del 2014

Srta.  
Janina Silvana Ortiz Pasaca  
*TESISTA DE LA CARRERA DE INGENIERÍA EN SISTEMAS*

De mi consideración:

Mediante la presente me permito saludar a su persona para indicarle que los nuevos datos en la Evaluación de Desempeño Docente del periodo 2014, no fueron otorgados debido a un cambio de administración en la comisión de evaluación interna, por lo cual se indica que la petición de los mismos se lo debe extender hacia el vicerrectorado de la universidad o hasta nuevo aviso a quien corresponda tal responsabilidad y coordinación.

Por la comprensión brindada le extiendo mis agradecimientos.

Atentamente.



Ing. Milton Labanda  
Técnico de Desarrollo de Software  
Unidad de Telecomunicaciones e Información

## **ANEXO 10: INFORME EJECUTIVO**

### **RESUMEN EJECUTIVO**

#### **Introducción:**

Los procesos que se llevan a cabo en instituciones públicas como privadas generan grandes cantidades de información, la Universidad Nacional de Loja en su proceso de Evaluación de Desempeño Docente no es la excepción. Actualmente dicha evaluación arroja resultados que evidencian la calificación de un docente en un rango de Insatisfactorio a Muy Satisfactorio, información útil pero no suficiente si se tiene presente que el proceso almacena más información, por ejemplo las opiniones de los estudiantes hacia determinado estudiante, en una asignatura, dentro de una carrera y en determinado periodo de evaluación así también componentes de evaluación.

Todo este conglomerado de generación de información procesada mediante técnicas inteligentes en este caso técnicas de minería de datos genera conocimiento que aporta a la toma de decisiones dentro del ambiente universitario.

#### **Problemática:**

El presente trabajo tiene como finalidad abarcar el proceso de evaluación de Desempeño Docente mediante un análisis con técnicas de Minería de Datos (MD), ya que en la actualidad dicho proceso, arroja resultados finales, que miden el desempeño en base a las puntuaciones obtenidas por el instrumento de evaluación ya sea encuesta o cuestionario, pero la misma no permite evidenciar donde existen realmente docentes que presenten criterios de interés u opiniones de los alumnos que aporten en plan de mejora hacia los docentes evaluados. Tampoco se permite conocer la evolución de cátedra que ha tenido un docente a través de los años, o los factores que influyen al momento que un estudiante evalúa a un docente.

Dentro de este contexto se ha creído necesario la aplicación de técnicas MD en la identificación de patrones, obtención de conocimiento y generación de modelos que contribuyan a la toma de decisiones tanto académicas como administrativas.

**Recursos:**

La realización de la presente investigación se la llevó a cabo en un escenario real como es la Universidad Nacional de Loja partiendo de los siguientes recursos:

- El Área de la Energía las Industrias y los Recursos Naturales No Renovables
- 3 Períodos de Evaluación de desempeño docente: 2011, 2012, 2013
- 3 Encuestas de Evaluación: 2 Satisfacción Estudiantil, 1 Evaluación de Desempeño
- Con 557 411 registros
- 517 991 evaluaciones
- 39 420 opiniones-estudiantes
- 11 506 estudiantes por asignatura
- 545 docentes por carreras y asignaturas

**Objetivos del Negocio:**

Los objetivos del Negocio permitieron al analista entender cuál es la necesidad que se deseó abarcar los mismos se menciona a continuación:

- Obtener los docentes que presente criterios de interés (opiniones de estudiantes) y buen desempeño docente, que se puedan dar dentro del proceso.
- Observar la evolución de cátedra que han tenido los docentes en los periodos académicos tomados para el estudio.
- Determinar los patrones de comportamiento de estudiantes que evalúan a sus docentes luego de obtener una nota en determinada asignatura.

**Resultados y Aporte:**

Dado que el presente proyecto es conveniente en el apoyo a la toma de decisiones a continuación se resume los resultados obtenidos y el aporte que los mismos pueden generar:

Criterios de Interés existentes en el proceso de evaluación docente: la minería de datos permitió detectar los criterios de interés o buen rendimiento que han tenido los docentes dentro de los periodos de estudio (ver figura 1)

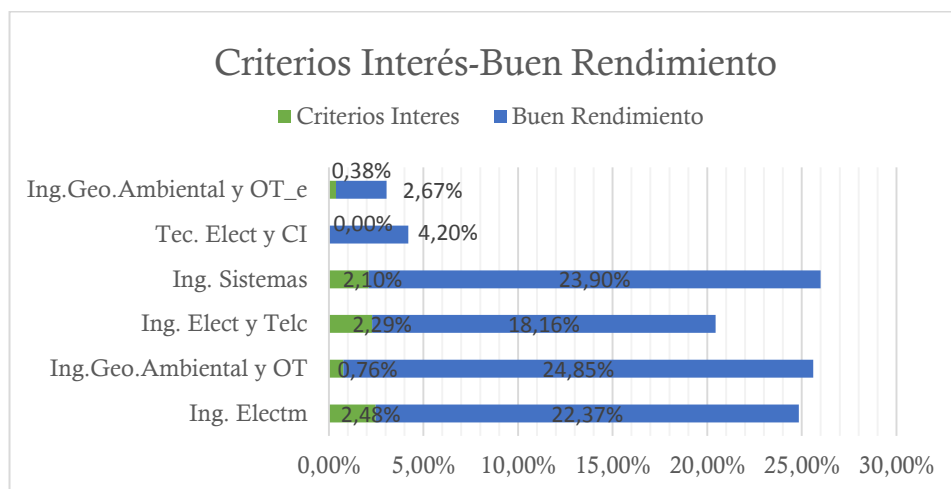


Figura 88: Criterios Interés-Buen Rendimiento, docentes del AEIRNRR.

El gráfico anterior muestra la existencia de docentes con criterios de interés y buen rendimiento, se puede evidenciar, una vez más que en todas las carreras existen docentes con un buen rendimiento en base a las calificaciones de los estudiantes y que existe un mínimo de docentes que presentan criterios de interés por cada carrera, cabe señalar que el porcentaje fue obtenido de la totalidad de docentes analizados. Es decir las carreras de Ing. En Geología Ambiental presencial y especial, Tecnología en Electricidad y Control Industrial presentan el porcentaje más bajo en docentes con criterios de interés con un 0.76%, 0.38% y 0.00%, respectivamente, y la carreras con más docentes con criterios de interés es la carrera de Ing. En Electromecánica, Sistemas y Electrónica y Telecomunicaciones. La mayor cantidad de docentes con un Buen Rendimiento, se evidencian dentro de la carrera de Ing. Geología Ambiental y Ordenamiento Territorial.

Luego de la obtención de los docentes que presentan criterios de interés y buen rendimiento dentro de la evaluación se procedió a obtener las opiniones de los estudiantes en base a dichos aspectos, tomando como referencia el número de identificación de cada docente. Cabe destacar que por motivo de confidencialidad se omiten las cédulas que corresponden a cada docente. A continuación se presenta un fragmento de las opiniones que mayor ocurrencia presentan dentro la evaluación (ver tabla I, II):

Palabra	Ocurrencia	Contexto
enseñanza	143	<ul style="list-style-type: none"> <li>• Que el docente mejore en algunos aspectos relacionados con la teoría y enseñanza.</li> <li>• Se debe explicar mejor los temas de la materia, que se escuche cuando un alumno comunica que no entiende las explicaciones, mejorar la enseñanza.</li> <li>• La enseñanza que sigue el docente no va de acuerdo a los intereses y objetivos de los estudiantes.</li> <li>• Enseñanza. Que cambie la forma de enseñar falta planificación</li> </ul>
mejorar	140	<ul style="list-style-type: none"> <li>• Mejorar su pedagogía, enseñanza, capacidad profesional.</li> <li>• Mejorar la comunicación docente-alumno.</li> </ul>
falta	103	<ul style="list-style-type: none"> <li>• Falta de preparación de las clases y falta del cumplimiento del sílabo.</li> <li>• Falta de comprensión docente-alumno, que el proceso de enseñanza sea más estricto para que los trabajos se realicen con más claridad.</li> <li>• Falta de planificación.</li> <li>• Falta de explicación</li> <li>• Falta de Interacción docente-alumno.</li> <li>• Falta de dominio en los contenidos.</li> <li>• Falta metodología de enseñanza</li> </ul>
conocimientos	87	<ul style="list-style-type: none"> <li>• Que se implemente una nueva metodología para cimentar de una mejor manera los conocimientos.</li> <li>• Explicar los contenidos elegidos, mayor dominio de los conocimientos.</li> <li>• Prepararse mejor para que se pueda explicar correctamente los conocimientos.</li> <li>• Los conocimientos adquiridos, son pocos en relación a los que se esperaba, no se trata temas a profundidad.</li> </ul>

*Tabla 1: Criterios Interés- Mayor Número De Ocurrencias*

Palabra	Ocurrencia	Contexto
bien	385	<ul style="list-style-type: none"> <li>• Bien: promueve la investigación en los alumnos.</li> <li>• Bien: siga cumpliendo de la misma forma su trabajo.</li> <li>• Bien: muy buenos logros, enseña a ser una mejor persona.</li> <li>• Bien: escucha al estudiante.</li> <li>• Bien: ha sido muy ético, se comprendió la totalidad de la materia.</li> </ul>
buenos	117	<ul style="list-style-type: none"> <li>• Presenta buenos valores éticos y morales.</li> <li>• Buenos valores, buen profesor éticamente un gran profesional.</li> <li>• Buenos conocimientos adquiridos a través del docente.</li> <li>• Buenos logros, gracias al docente.</li> <li>• Buenos logros académicos con la metodología aplicada</li> </ul>
bueno	176	<ul style="list-style-type: none"> <li>• Buena relación con los estudiantes.</li> <li>• Buena metodología de enseñanza, las clases son dinámicas, existe conocimientos.</li> <li>• Buena ética profesional.</li> <li>• Buena persona, califica adecuadamente.</li> <li>• Buena persona tanto personal como profesional.</li> </ul>

excelente	194	<ul style="list-style-type: none"> <li>• Excelente docente, persona, profesor, profesional se preocupa por los estudiantes, responsable en lo que realiza.</li> <li>• Excelente comportamiento es un gran aporte para la carrera.</li> </ul>
-----------	-----	--

Tabla II: Opiniones: Buen Rendimiento- Mayor Número De Ocurrencias

### Evolución de cátedra que han tenido los docentes en los diferentes periodos académicos:

La evolución de cátedra que han tenido los docentes en todas las carreras del AEIRNNR presenta una estadística muy grande, las mismas que evidencian el nivel de preparación y profesionalidad existente en cada una de ellas (ver Figura 2).

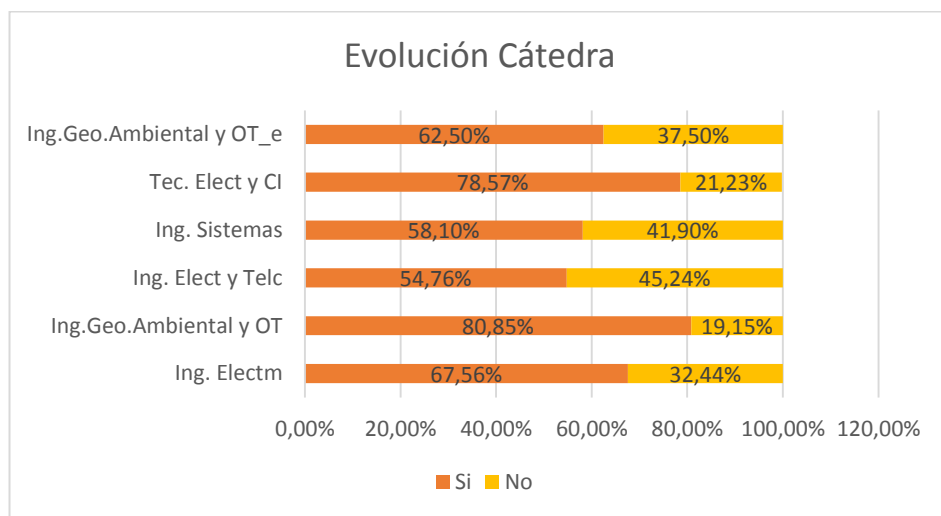


Figura 2: Docentes que No presentan una evolución de cátedra

En las figuras se muestra la cantidad de docentes que presentan una evolución de cátedra por carrera, como aquellos que no la poseen. Se puede observar que todas las carreras tienen un buen número de docentes con evolución de cátedra, siendo un aspecto fundamental en la calidad de la educación superior. En la obtención de una buena calidad en la educación es necesario contar con docentes que a medida pase el tiempo vayan evolucionando y adquiriendo más experiencia tanto profesional como personal, esto se cumple en su mayoría dentro del AEIRNNR.

Pero cabe resaltar que las carreras con más docentes con evolución de cátedra son las carreras de Geología Ambiental y Ordenamiento Territorial, seguido de la carrera de Tecnología en Electricidad y Control Industrial con un 80.85% y 78.57% respectivamente. Las carreras con menos docentes sin dicha evolución son las carreras de Ingeniería en Electrónica y Telecomunicaciones e Ingeniería en Sistemas, con un 45.24% y 41.90% del total de docentes de dichas carreras.

Este análisis refuerza también refuerza el factor crítico estudiando anteriormente que son los criterios de interés ya que las carreras que presentan menos evolución de cátedra anteriormente también presentaron más criterios de interés. Lo descrito se puede evidenciar en la Sección Resultados: Validación del Modelo y Análisis de Resultados.

**Grado de Influencia que tienen las Notas de los Estudiantes en la Calificación de Desempeño Docente.**

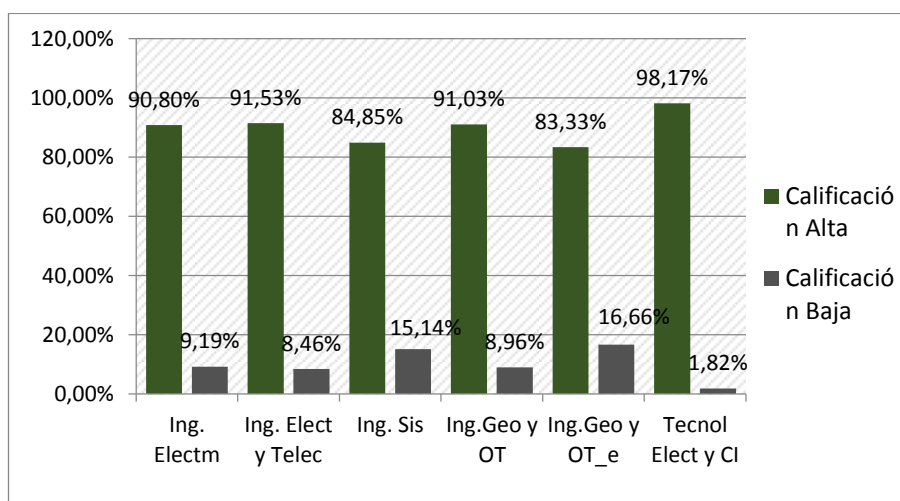


Figura 3: Influencia de Notas en Calificaciones AEIRNNR

Se puede interpretar además que la mayor parte de docentes que posee calificaciones altas es la carrera de Tecnología en Electricidad y Control Industrial con un 98.17%, seguido de Ingeniería en Electrónica y Telecomunicaciones con 91.53%.

Las carreras de Ingeniería en Geología Ambiental y Ordenamiento Territorial y la Carrera de Ingeniería en Sistemas poseen más calificaciones bajas con un 15.14%, 16.66% respectivamente (ver figura 4)



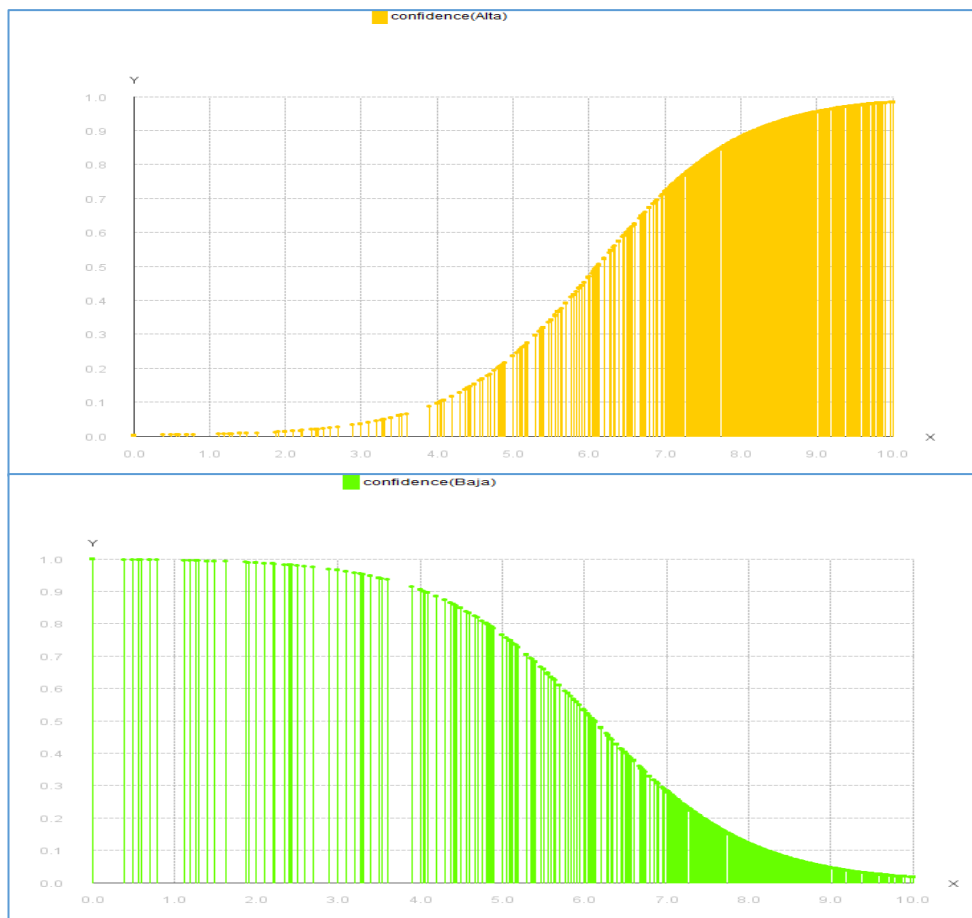


Figura 4: Influencia de Notas en Calificaciones Docentes

En el gráfico se puede reforzar el análisis realizado con los diferentes algoritmos y modelos generados, es decir a mayor nota de un estudiante, más es la calificación de un docente en la evaluación de desempeño (Ver Sección Resultados: Validación del Modelo y Análisis de Resultados). Esto ocurre en todas las carreras dando un patrón de comportamiento a mayor nota estudiante, entonces alta calificación docente y a menor nota estudiante, entonces baja calificación docente. Por lo tanto también se realizó un análisis de las preguntas del cuestionario para verificar si las mismas son procedentes o no al criterio de evaluación de un estudiante (Ver Anexo 11). Con los resultados obtenidos a continuación se detallan los aportes en los cuales la presente investigación aporta:

**Aporte1:** las opiniones vertidas por los estudiantes en el proceso de evaluación de desempeño servirá al apoyo a la toma de decisiones en cuanto a la elaboración de un plan de mejoras tanto para docentes que requieran reforzar ciertos aspectos, como pueden ser tomadas en cuenta para docentes que presenten un buen rendimiento en cuanto a la aplicación de incentivos y reconocimientos de cátedra.

**Aporte 2:** la evolución de cátedra es un aspecto fundamental en la calidad de la educación superior, al ser tomados en cuenta se pueden implementar o diseñar cursos específicos para docentes que no posean dicha evolución, los cursos diseñados y el cumplimiento de los mismos pueden ser ejecutados por carreras.

**Aporte 3:** con la obtención de estos patrones se puede aportar a la toma de decisiones dentro del ambiente educativo en cuanto a la evaluación de desempeño docente, para promover y concienciar a los estudiantes a un proceso serio que debe ser evaluado de acuerdo a como se desempeñó el docente en su espacio de cátedra, y no por las notas obtenidas en determinado módulo.

### **Conclusión y Beneficios:**

La aplicación de técnicas de minería de datos en la evaluación de desempeño docente permitió obtener patrones de comportamiento, o factores de interés que a simple vista o con la ejecución solo de consultas sql no sería fácil de obtener. Los modelos generados, algoritmos aplicados y herramientas utilizadas permitieron llevar a cabo la presente investigación.

Se sabe que todo proceso ya sea académico o empresarial tiene sus fortalezas y debilidades, la evaluación de desempeño docente no es la excepción. La presente investigación ayudaría a:

- La designación de los docentes o en la renovación de sus cargos en base a su evolución de cátedra o buen rendimiento presentado.
- Atacar las falencias existentes en el espacio de cátedra de docentes que realmente necesiten reforzar ciertos aspectos profesionales en el proceso enseñanza-aprendizaje.
- Estimar el grado de influencia que tendrán las notas estudiantiles hacia la evaluación docente de tal manera que se proponga un constante rediseño de los instrumentos de evaluación como encuestas o cuestionarios en base a los reglamentos establecidos por la LOES.

## ANEXO 10: Artículo Científico

# Análisis del proceso de Evaluación al Desempeño Docente, aplicando Técnicas de Minería de Datos

Ortiz-Pasaca, Janina\* ; Chamba-Eras, Luis\*

\*Universidad Nacional de Loja, Carrera de Ingeniería en Sistemas,  
Loja, Ecuador, e-mail: {jsortizp, lachamba}@unl.edu.ec

**Abstract**— This paper presents the application of data mining (DM) techniques in the evaluation of teaching performance, the main objective is to obtain patterns within the cluster of data available. The research is oriented towards achieving the criteria of interest, useful to attack the weaknesses of academic space teachers really need to strengthen certain professional aspects in the teaching-learning process. Examine the development of chair that had a teacher through the years and also see the influence of student notes to the teacher evaluation. Detailed, implementation of algorithms, techniques and validation of DM generated models, in a real scenario as in the Area of Energy, National University of Loja.

**Keywords**— *Data Mining, Algorithms, Evaluation, Interest Criteria, Evolution, Performance.*

### I. Introducción

La revolución tecnológica, sin duda alguna ha hecho posible que la información digitalizada sea fácil de capturar, almacenar, distribuir, y transmitir [1]. Es así que en la actualidad dicha revolución provoca que cada vez exista mayor dependencia en el uso sistemas de información y por ende almacenamiento de datos en todas las empresas o entidades públicas o privadas. Las instituciones educativas son entidades que por lo general, almacenan gran cantidad de información que permanece estancada, en Bases de Datos, siendo útil únicamente en la generación de reportes o consultas. Se sabe claramente que la información es el activo más importante de una empresa o institución por ende la producción de la misma abarca la generación de estrategias, incremento de ganancias, automatización de procesos, pero sobre todo apoyo a la toma de decisiones [2].

Dentro de este gran conglomerado de generación de información académica, se encuentra el proceso de Evaluación de Desempeño Docente, que juega un papel muy importante en la calidad de la educación superior en Ecuador.

La calidad en la educación superior es un factor primordial en el proceso enseñanza-aprendizaje y en cuanto a procesos educativos, se puede entrever que el objetivo principal de la calidad es un buen rendimiento o un mejoramiento continuo en base a los recursos académicos disponibles.

La educación superior en el Ecuador ha marcado grandes cambios en los últimos años, una nueva corriente de transformación, que va

desde la creación de instituciones hasta el establecimiento de consejos como: el Consejo de Educación Superior (CES) y el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES), que aseguran primordialmente la calidad en la educación al cumplir con estándares establecidas por el Estado [3, 4]. Es así que dentro de esta gran transformación se encuentra el proceso de evaluación al desempeño docente, que en este caso busca la calidad y excelencia de cátedra.

Se puede referir a la evaluación del desempeño docente, como un conjunto de acciones organizadas en base a las intervenciones e interacciones pedagógicas, entre el conocimiento científico y el conocimiento académico, lo cual permite impulsar acciones didácticas pedagógicas que fortalezcan los procesos de aprendizaje de los estudiantes, y el mejoramiento de la formación docente, y a su vez su desarrollo profesional [5, 6].

El proceso de evaluación al desempeño docente en la actualidad, tiene resultados finales, que miden el desempeño en base a las puntuaciones obtenidas por el instrumento de evaluación ya sea encuesta o cuestionario, pero la misma no permite evidenciar donde existen

realmente criterios de interés u opiniones de los alumnos que aporten en plan de mejora hacia los docentes evaluados. Tampoco se permite conocer la evolución de cátedra que ha tenido un docente a través de los años, o los factores que influyen al momento que un estudiante evalúa a un docente. Herramientas o técnicas que proporcionen un análisis en este tipo de procesos educativos, son importantes [7-10] para conocer patrones en la evaluación de desempeño al docente.

Por tal motivo actualmente es necesario aplicar técnicas de Minería de Datos (MD) que permita el análisis en el proceso de evaluación de desempeño docente, para en base a la presente investigación se planteen estrategias de acción necesarias para la capacitación, actualización profesional y desarrollo personal, de los docentes universitarios en el contexto de la nueva Ley de Educación Superior y las necesidades institucionales.

Para el lector la organización del trabajo es la siguiente: I INTRODUCCIÓN, presenta el objeto de estudio, problemática y justificación. II ESTADO DEL ARTE, detalla conceptos, trabajos relacionados, técnicas y herramientas utilizadas. III METODOLOGÍA se describe el modelo y validación de la minería de datos. IV RESULTADOS y DISCUSIÓN se detalla un análisis del experimento realizado, así como la discusión de los resultados obtenidos. V CONCLUSIONES, establece los logros alcanzados y las líneas futuras que se generó tras la culminación del trabajo de investigación.

## I. Estado del Arte

La Minería de Datos de aquí en adelante (MD) es un proceso que consiste en analizar los datos desde diferentes perspectivas, en base a la extracción de los mismos y convirtiéndolos en información útil. Pero además permite descubrir las relaciones en bases de datos que pueden identificar comportamientos que no son muy evidentes, es decir permite la identificación de patrones ocultos y generación de conocimiento [11].

Dentro de la educación la MD, se ha convertido en un área de estudio con grandes expectativas en cuanto a la resolución de problemas y generación de conocimiento en base a los datos disponibles.

Se puede citar varias investigaciones en donde se recurre a la MD como herramienta de solución e innovación en el ámbito educativo [8-10]. Entre estas se puede citar a la “Minería de Datos en Encuestas de Profesores al Fin de Semestre de La Facultad de Ingeniería, UNAM” [8], cuyo objetivo fue obtener reglas o patrones

sobre el comportamiento y desempeño docente. Estas reglas o patrones tienen la finalidad de brindar apoyo a las divisiones y departamentos que conforman el cuerpo académico, con el objetivo de mejorar el desempeño de los profesores, así como obtener una mejor calidad en la enseñanza-aprendizaje. De la misma manera una investigación acerca de la “Obtención de Patrones y Reglas en el Proceso Académico Utilizando Técnicas de Minería de Datos” [9], muestra un estudio para determinar el vínculo que existe entre el origen y procedencia social de los estudiantes con sus resultados académicos, mediante la aplicación de técnicas de agrupación y reglas de asociación de MD. El estudio en sí se orientó a la predicción de las notas de las asignaturas del primer año de estudiantes basado en las relaciones que se establecen entre el nivel académico de los padres, tipo de centro de procedencia, provincia y resultados académicos.

### A. Técnicas de Minería de Datos.

Existen dos técnicas de MD: descriptivas y predictivas. Las técnicas descriptivas permiten la identificación de comportamientos, tendencias o descubrir características significativas en los datos, mientras que las técnicas predictivas exponen o resumen las variables de interés o estiman valores futuros [8].

La Tabla I [8] presenta la clasificación de las técnicas de MD, con sus respectivos algoritmos.

TABLA I. TÉCNICAS DE MINERÍA DE DATOS

Técnicas		Algoritmos de Minería de Datos		
Predictivas o Supervisadas	Clasificación	Árboles de Decisión	Redes Bayesianas	Redes Neuronales
	Regresión	Regresión Logística	Algoritmos Genéticos	Generación de Reglas
Descriptivas o No supervisadas	Agrupación	Mapas de Kohonen.	K medias	
	Asociación	Reglas de Asociación		

En base a los trabajos relacionados con la investigación, se pudo recopilar bibliográficamente diversas técnicas de MD las cuales dieron las bases para la selección de las técnicas apropiadas para el escenario de investigación. Las técnicas escogidas para el estudio son: árboles de decisión, técnicas de regresión, generación de reglas y técnicas bayesianas [7-10]. Es necesario acotar, que se puede aplicar varias técnicas de MD para la resolución de un problema, y con experimentos se procede a comparar resultados del o los modelos generados.

## B. Herramientas de Minería de Datos.

Para la selección de las herramientas se realizó un análisis comparativo que permitió determinar que Weka y RapidMiner son las herramientas que ofrecen más ventajas para abordar el escenario de estudio, ya que ambas herramientas se complementan computacionalmente. Se seleccionó a RapidMiner como la herramienta principal por la variedad de características que ofrece en cuanto al proceso de minería, y a la vez incluyen muchas funcionalidades de la herramienta Weka. Weka sirvió de apoyo en la aplicación de algoritmos que en RapidMiner no están bien definidos [12,13].

## III. Metodología

La información para el estudio se la obtuvo por medio del Sistema de Evaluación de Desempeño Docente (EDD) y del Sistema de Gestión Académica de la Universidad Nacional de Loja, proporcionada por la Unidad de Telecomunicaciones e Información, datos en físico y a través de un Servicio Web (Web Services). Para la investigación se consideró los resultados obtenidos en el proceso de evaluación al docente durante los periodos académicos 2011, 2012 y 2013, en las 4 carreras de grado del Área de la Energía de las Industrias y los Recursos Naturales No Renovables.

La presente investigación utilizó una metodología que especifica los pasos y procesos que se debe seguir para la culminación de cualquier proyecto de MD, de manera satisfactoria, la cual es Cross Industry Standard Process for Data Mining (CRISP-DM) [14,15], muy utilizada en el campo académico y empresarial.

CRISP-DM constituye la guía de referencia más utilizada en el desarrollo de proyectos de explotación de datos, en el proyecto se ejecutaron las siguientes fases: comprensión del negocio o problema, comprensión de los datos, preparación de los datos, modelado, evaluación, validación, e implantación [14].

### A. Fase uno: Comprensión del Negocio

Fase que permitió determinar y definir claramente los problemas a los que se va a dar solución, esto permitió recolectar los datos correctos e interpretar correctamente los resultados. También se analizó los recursos disponibles para el desarrollo del análisis como recursos hardware, software, materiales-servicios y talento humano, pero además se definió los criterios de éxito del proyecto desde un punto de vista del negocio los cuales se detallan a continuación:

- La identificación de patrones dentro del proceso de evaluación es tomado como un criterio de éxito en el desarrollo del presente trabajo investigativo, tomando como base los datos de la EDD del AEIRNNR de la Universidad Nacional de Loja.
- Las notas también se utilizan, para detectar el comportamiento de los estudiantes al momento de evaluar, y la evolución de cátedra que un docente tiene en base a las evaluaciones, fueron criterios de éxito dentro de la minería de datos en el presente proyecto.

### B. Fase Dos: Comprensión de los Datos

La segunda fase, corresponden a la comprensión de los datos, la recolección inicial de los mismos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes.

La comprensión de los datos abarca la creación de funciones, vistas y consultas que permitan a su vez el análisis de la información y registros que contiene la BD disponible. La exploración de la base de datos presenta información relevante como: el número de estudiantes dentro de los periodos de evaluación abordados (ver figura 1), la carrera de Ingeniería en Sistemas es la carrera que más estudiantes han participado en el proceso de EDD ya que presenta en los periodos académicos un total de 729 estudiantes. Parcialmente se tiene un total de 1833 en los 3 periodos académicos, a ello se le agrega un subtotal de 77 estudiantes que pertenecen a las carreras especiales dando un resultado final de 1910 estudiantes.

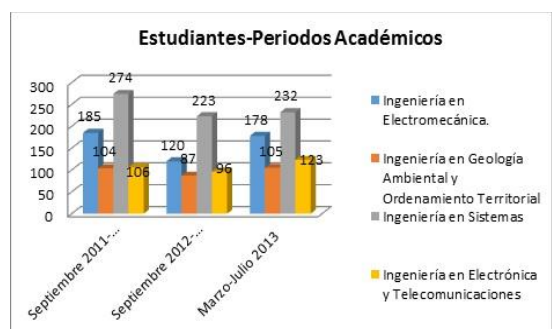


Fig. 1. Número estudiantes- Periodos Académicos.

En cuanto a las evaluaciones cumplidas por los docentes en los periodos de estudio, se puede concluir, que la carrera de Ingeniería en Electromecánica tiene un 56% de docentes que completan los 3 periodos, un 11% que completan 2 periodos y un 33% que tienen 1 periodo de evaluación. El total de docentes es de 100 en los 3 periodos, se acota que este total difiere del total de docentes por periodo de evaluación.

### C. Fase Tres: Preparación de los Datos

La preparación de datos incluye las tareas generales de selección de datos a los que se aplicó una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. Dentro de esta fase también se construyó las estructuras necesarias para la aplicación de los algoritmos seleccionados y generación de los modelos. Las estructuras constan de la selección, generación y discretización de los atributos (ver tabla II):

TABLA II. ATRIBUTOS DE LAS ESTRUCTURAS DE MD.

<b>cedula_docent, cedula_estudiante</b>
Atributo de tipo varchar, continuo, y corresponde a los números de identificación de cada docente del AEIRNNR. Esto con el fin de agrupar a los mismos para posteriormente conocer el rendimiento y la evolución de cátedra que tienen.
<b>carrera:</b>
Atributo de tipo varchar, discreto, útil para generar información acerca del desarrollando que tienen los docentes dentro de determinada carrera, las mismas que pueden ser: Ing. en Elect, Ing en Geol y OT, Ing. en Sist, Ing. Elect y Tele, Ing en Geol y OT_e, Tec_ Elec_CI
<b>asignatura:</b>
De tipo varchar, continuo es un atributo muy importante dentro del análisis, motivo por el cual se toman todas las asignaturas impartidas dentro del AEIRNNR, para posteriormente conocer patrones de comportamiento.
<b>periodo_acad_id:</b>
Atributo de tipo integer, discreto, útil en la clasificación de tiempos correspondientes a cada proceso de evaluación. Puede tomar valores binarios en diferentes estructuras
<b>calf_4_3, calf_2_1:</b>
De tipo integer, atributo que permite conocer los ítems más altos y más bajos como respuesta de la evaluación de desempeño docente.
<b>cpf_4_3, cpg_4_3, pv_4_3:</b>
De tipo integer, atributo que permite conocer los ítems más altos como respuesta de la evaluación de desempeño docente en cada componente de evaluación.
<b>cpf_2_1, cpg_2_1, pv_2_1:</b>
De tipo integer, atributo que permite conocer los ítems más bajos como respuesta de la evaluación de desempeño docente en cada componente de evaluación.
<b>porcentaje:</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación obtenido por un docente, en una asignatura, calculado de la multiplicación entre calf_4_3 por 100 dividido entre el total de ítems calificados o resultado.
<b>evaluacion:</b>
De tipo varchar, atributo que discretiza el porcentaje de evaluación obtenido por un docente en: Criter-Int: 0-69 %, Buen-Rend: 70-100 %
<b>resul_cpf, resul_cpg, resul_pv</b>

De tipo integer, continuo, atributo que muestra el total de ítems en cada uno de los componentes de evaluación, por ejemplo cpf_4_3 y cpf_2_1, total de capacidad profesional.
<b>porc_cpf, porc_cpg, porc_pv</b>
De tipo integer, continuo, atributo que muestra el porcentaje de calificación en cada uno de los componentes de evaluación, por ejemplo en capacidad profesional obtenida por un docente, en una asignatura, calculado de la multiplicación entre cpf_4_3 por 100, dividido para resul_cpf. Se discretiza en: regular: 0-69 %, buena: 70- 80 %, excelente: 81-100 %
<b>catedra</b>
De tipo varchar, atributo que permite conocer la calificación obtenida por un docente dentro de un periodo académico basado en los ítems más bajos de calificación (2 y1 ) calculado de la multiplicación entre calf_2_1 por 100, dividido para el total de ítems calificados o resultado, obteniendo un porcentaje del 0 al 11 % y del 12 al 100 %. Se discretiza en: <b>P</b> : positiva, <b>N</b> : negativa
<b>catedra_cpf, catedra_cpg, catedra_pv</b>
De tipo varchar, atributo que permite conocer la calificación obtenida por un docente dentro de un periodo académico por cada componente de evaluación puede ser: <b>N</b> : negativa, <b>P</b> :positiva
<b>catedra_p_1, catedra_p_2, catedra_p_3</b>
De tipo varchar, atributo que permite conocer la calificación de un docente por periodo académico basado en el atributo catedra, puede ser: <b>N</b> : negativa, <b>P</b> :positiva, <b>A</b> : ausencia en determinado periodo académico
<b>resultado, total_calf:</b>
De tipo integer, continuo, atributo que muestra el total de calf_4_3 y calf_2_1, total de preguntas en un cuestionario.
<b>calif_cualitativa:</b>
De tipo varchar, discreto, atributo que muestra el rendimiento de un docente, basado en los resultados (calf 4 y 3) de las evaluaciones en una materia determinada. El mismo que pueden ser: Muy alto: 81-100 %, Alto: 70-80 %, Bajo: 66-69 %, Muy bajo: 0-65 %
<b>evol_catedra</b>
Atributo de tipo discreto, denota la evolución de cátedra que ha tenido un docente en los periodos académicos de estudio, clasificado en: Si, No
<b>calif_cualitativa:</b>
De tipo varchar, discreto, atributo que muestra el rendimiento de un docente, basado en los resultados (item 4 y 3) de las evaluaciones en una materia determinada. El mismo que puede ser: Baja: 0-70 %, Alta: 71-100 %
<b>opiniones:</b>
De tipo varchar, continuo, atributo que permite la recopilación de los criterios más comunes generados del proceso de evaluación docente.
<b>nota:</b>
De tipo double, continuo, categoriza el promedio de la nota obtenida por un estudiante en determinada asignatura: Mala: 0 – 6.9, Regular: 7 – 7.5, Buena: 7.6 – 8.9, Excelente: 9 – 10

#### D. Fase Cuatro: Modelado

En esta fase de modelado se seleccionó las técnicas de minería apropiadas para el conjunto de datos provistos, como se mencionó, se puede aplicar varias técnicas para el mismo problema de minería.

Las técnicas seleccionadas para identificar la existencia de criterios de interés dentro de una estructura, fueron los árboles de decisión con los algoritmos ID3, CHAID y C4.5, además de la generación de reglas a través del algoritmo NNge y PART. Los resultados en el rendimiento de cada algoritmo se describen en la tabla III.

TABLA III. RENDIMIENTO DE ALG.CRITERIOS INTERÉS.

Modelo	Evaluación	Valores correctamente clasificados	Valores incorrectos en la clasificación
ID3	Desempeño Modelo	100 %	0.00 %
	Validación Cruzada	97.25 %	2.75 %
CHAID	Desempeño Modelo	99.82 %	0.18 %
	Validación Cruzada	98.90 %	1.10%
C4.5	Desempeño Modelo	96.70 %	3.30 %
	Validación Cruzada	94.32 %	5.68 %
NNge	Desempeño Modelo	100 %	0.00 %
	Validación Cruzada	98.89 %	1.10 %
PART	Desempeño Modelo	99.82 %	0.18 %
	Validación Cruzada	99.27 %	0.73 %

Todos los algoritmos aplicados obtuvieron una buena clasificación, excepto el algoritmo ID3, pero en comparación con los demás algoritmos. Por lo tanto se lo sigue considerando porque supera una clasificación en la validación cruzada del 90%. La obtención de las opiniones de los estudiantes en el proceso de evaluación se lo realizó mediante las herramientas Textalyser y Many Eyes.

Los algoritmos aplicados en la evolución de cátedra de los docentes a través de los años fueron: Naive Bayes, ID3, REPTree y NNge. Los resultados en el rendimiento de cada algoritmo se describen en la tabla IV.

TABLA IV. RENDIMIENTO ALG. EVOLUCIÓN DE CÁTEDRA.

Modelo	Evaluación	Valores correctamente e clasificados	Valores incorrectos en la clasificación
Naive Bayes	Desempeño o Modelo	93.05 %	6.95 %
	Validación Cruzada	93.05 %	6.95 %
ID3	Desempeño o Modelo	97.30 %	2.70 %
	Validación Cruzada	92.28 %	7.72 %
REPTree	Desempeño o Modelo	96.14 %	3.86 %
	Validación Cruzada	94.21 %	5.79 %
NNge	Desempeño o Modelo	96.91 %	3.08 %
	Validación Cruzada	94.21 %	5.79 %

En la tabla IV se observa que todos los algoritmos tienen una clasificación tanto para el desempeño del modelo, como en la validación cruzada mayor al 90 %.

En la tabla V se presentan los algoritmos aplicados al grado de influencia que tienen las notas de los estudiantes en las calificaciones de desempeño docente. Los algoritmos aplicados fueron: regresión logística, ID3 y C4.5.

TABLA V. RENDIMIENTO DE ALG. INFLUENCIA NOTAS .

Modelo	Evaluación	Valores correctamente clasificados	Valores incorrectos en la clasificación
Regresión Logística	Desempeño Modelo	88.77 %	11.23 %
	Validación Cruzada	88.74 %	11.26 %
ID3	Desempeño Modelo	88.96 %	11.04 %
	Validación Cruzada	88.50 %	11.50 %
C4.5	Desempeño Modelo	88.68 %	11.32 %
	Validación Cruzada	88.73%	11.27 %

En la tabla V se puede observar que existe un alto grado de precisión en los modelos de regresión logística y el árbol de clasificación C4.5, a diferencia del árbol de decisión ID3 si bien es cierto baja el grado de precisión, sus resultados y condiciones pueden ser tomadas en cuenta en base a los modelos anteriores. A diferencia de los modelos anteriores, se puede

afirmar que la calidad de los datos influye en la generación de los modelos en este caso la existencia de valores nulos [16].

### E. Fase Cinco: Evaluación

Para la evaluación de los modelos se hizo un análisis comparativo que permita determinar el grado de acierto para las predicciones y clasificaciones realizadas, se tomaron como medidas de evaluación el error absoluto y el error relativo, kappa y error cuadrático medio.

**Criterios Interés - Buen Rendimiento:** se puede concluir que el algoritmo que mejor se desempeña con los datos disponibles es el algoritmo ID3, CHAID, NNge y W-PART, sin descartar los resultados arrojados por el algoritmo C4.5 ya que tiene una precisión mayor del 90 %. En conclusión los algoritmos aplicados obtuvieron buenos resultados y útiles dentro del análisis propuesto, descartando los 2 últimos debido a la poca generación de reglas (ver figura 3)[17].

**Evolución de Cátedra:** los algoritmos que mejor se desempeñaron con los datos disponibles es el algoritmo REPTree, NNge y Naive Bayes, sin descartar los resultados arrojados por el algoritmo ID3 ya que tiene una precisión mayor del 90 %. En conclusión los algoritmos aplicados obtuvieron buenos resultados y útiles dentro del análisis propuesto, descartando los REPTree debido a que poseía reglas complejas de interpretar (ver figura 4) [17].

**Influencia de notas-estudiantes:** se puede concluir que los algoritmos que mejor se desempeñan con los datos disponibles el algoritmo de Regresión Logística y el algoritmo ID3, sin embargo existen condiciones en el árbol de decisión C4.5 que se tomaron en cuenta dentro del análisis (ver figura 5).

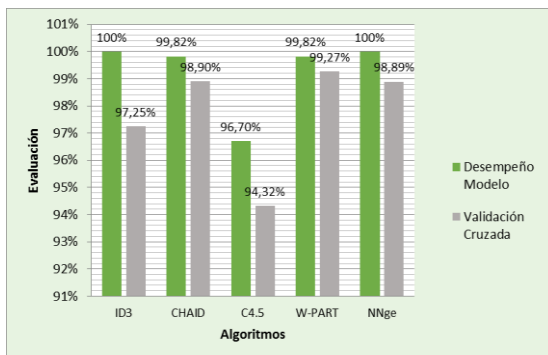


Fig. 3. Resultados obtenidos por cada algoritmo- Criterios

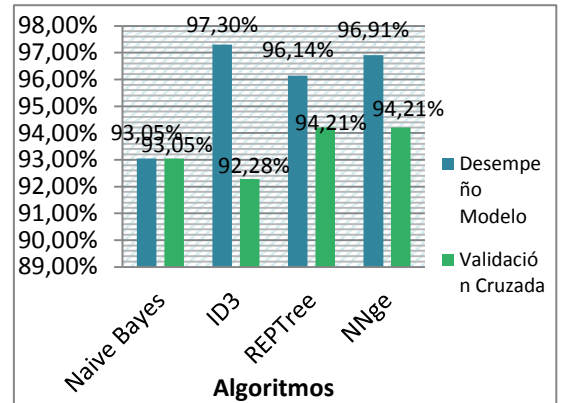


Fig. 4. Resultados obtenidos por cada algoritmo- Evolución Cátedra

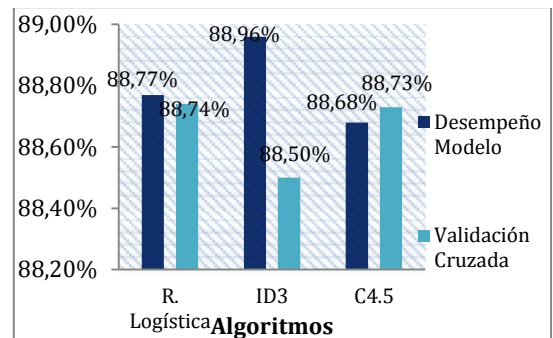


Fig. 5. Resultados obtenidos por cada algoritmo, Notas.

## IV. Resultados y Discusión

Al aplicar minería de datos dentro del proceso de evaluación de desempeño docente se pudo encontrar patrones que a simple vista no son muy evidentes dentro del conglomerado de datos a continuación se presenta una sección de los patrones encontrados:

**Criterios Interés - Buen Rendimiento:** las carreras con menos criterios de interés en el proceso de evaluación son las carreras de Ing. en Geología Ambiental y Ordenamiento Territorial, presencial y especial, seguido de la carrera de Tecnología en Electricidad y Control Industrial.

- El periodo donde existe la mayor cantidad de criterios de interés dentro de la evaluación docente es el periodo 2 (septiembre 2012) con un 38.09 % del total de criterios, y el mayor periodo favorable a buen rendimiento docente es el periodo 1 (septiembre 2011) con un 34.99 % del total.
- En el periodo académico 1 (septiembre 2011) existe un buen rendimiento debido a la capacidad profesional de cada docente la cual se clasifica en buena y excelente.
- En el periodo académico 1 (septiembre 2011) si la capacidad profesional y capacidad pedagógica es baja o regular pero existe una buena práctica de valores, entonces la carrera



de Ingeniería en Sistemas presenta un Buen Rendimiento en sus docentes.

**Obtención de Opiniones:** una vez identificado donde existen criterios de interés se procede a emplear *texminig* en la identificación de las demandas o refuerzos que los docentes deberían tener en cuenta, en el contexto del plan de mejoras, las mismas que son sugeridas por los estudiantes.

Las opiniones que se resaltan más cuando existen criterios de interés son las siguientes:

- El docente debe mejorar su metodología de enseñanza 40 % de las ocurrencias.
- El docente debe preparar clases 10 %.
- Se debe brindar más apoyo y supervisión en la elaboración de los trabajos 8 %.

**Evolución de Cátedra:** las carreras con más docentes con evolución de cátedra son los que pertenecen a las carreras Ingeniería en Geología y Ordenamiento Territorial presencial (especial), la carrera de Tecnología en Electricidad y Control Industrial y Electromecánica con un 78.72 %, 62.5 %, 78.57 %, 68.91 % respectivamente.

- Un docente que no ha tenido una evolución de cátedra por lo general tendrá una calificación de Muy baja o Baja, en una o varias evaluaciones de desempeño y negativo en alguno de los componentes evaluados de forma constante.
- Cuando un docente pertenece a la carrera de Ing. en Electrónica y Telecomunicaciones y en alguna evaluación posee una capacidad pedagógica o profesional negativa pero una constante práctica de valores positiva entonces si tiene una evolución de cátedra
- Los docentes que hayan obtenido en sus períodos de evaluación, primero una calificación Muy Baja, luego una calificación Alta y después se mantenga o suba a Muy Alta efectivamente tendrá una evolución de cátedra.

**Influencia de notas-estudiantes:** si un estudiante sobrepasa una calificación de 6 a 7, existe la probabilidad que un docente tenga una calificación alta.

- Si un estudiante tiene una calificación buena e incluso excelente, mayor será la probabilidad en ítems de calificación de 4 o 3 hacia un docente.
- En la mayoría de los casos si un estudiante tiene una nota menor a 6, menor será la probabilidad que un docente tenga una calificación alta en el proceso de evaluación en su desempeño.

En lo que concierne a la influencia de las notas de los estudiantes hacia las calificaciones

docentes se destaca un patrón de comportamiento: en donde se evidencia que a mayor nota de un estudiante en una asignatura, mayor probabilidad tiene un docente de obtener una buena calificación de desempeño y de igual manera a menor nota de un estudiante menor la probabilidad existe que un docente tenga una calificación de desempeño alta. Por lo tanto también se realizó un análisis de las preguntas del cuestionario para verificar si las mismas son procedentes o no al criterio de evaluación de un estudiante.

### B. Discusión

Se sabe que todo proceso ya sea académico o empresarial tiene sus fortalezas y debilidades, la evaluación de desempeño docente no es la excepción. La aplicación de técnicas de minería de datos sirvió para atacar las debilidades en cuanto a dicho proceso y en base a la investigación realizada proponer alternativas de mejora las mismas que servirían para:

- La designación de los docentes o en la renovación de sus cargos en base a su evolución de cátedra o buen rendimiento presentado.
- Atacar las falencias existentes en el espacio de cátedra de docentes que realmente necesiten reforzar ciertos aspectos profesionales en el proceso enseñanza-aprendizaje.
- Estimar el grado de influencia que tendrán las notas estudiantiles hacia la evaluación docente de tal manera que se proponga un constante rediseño de los instrumentos de evaluación como encuestas o cuestionarios en base a los reglamentos establecidos por la LOES.

### V. Conclusiones

La MD deja por delante un campo fértil y prometedor con muchos retos en investigación. Actualmente se está recurriendo a esta herramienta para la optimización de procesos empresariales, descubrimiento de patrones u obtención de conocimiento en bases de datos, apoyo la toma de decisiones e incremento de ganancias.

El análisis del proceso de evaluación docente empleando técnicas de MD puede ser mejorado al crear nuevas estructuras en base a nuevos atributos disponibles en el sistema de gestión académica que incidan en la generación de nuevos modelos.

El presente estudio da lugar a la generación de nuevas líneas de investigación, por ejemplo la implementación de *texminig* (minería de texto) a las opiniones de los estudiantes, docentes, directivos, de forma más detallada y

de esta manera obtener un nuevo modelo en base a los criterios identificados por la minería.

#### Referencias

- [1] S. Mitra and T. Acharya. Data mining: multimedia, soft computing and bioinformatics. John Wiley & Sons, 2003.C.
- [2] J. Arévalo. “Gestión de la Información, gestión de contenidos y conocimiento”. Universidad de Salamanca. 2007
- [3] R. Ramírez, A. Minteguiaga, Transformaciones En La Educación Superior Ecuatoriana: Antecedentes Y Perspectivas Futuras Como Consecuencia De La Nueva Constitución Política. Revista Ess, vol. 15, no. 1, pp 131-148, 2010.
- [4] R. Ramírez, Tercera Ola De Transformación De La Educación Superior En Ecuador. Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Investigación. 2013.
- [5] La Evaluación de la Actividad Docente Del Profesorado en la Universidad Pública De Navarra. Universidad Pública de Navarra. 2012.
- [6] H. Rizo, Evaluación Del Docente Universitario, Una Visión Institucional. Universidad Autónoma de Occidente. Revista Iberoamericana de Educación (ISSN: 1681-5653).
- [7] L. Jing, Data mining and its applications in higher education. New directions for institutional research. 2002. Vol.2002, no 113, p.7-30
- [8] A. Román, Minería de Datos en Encuestas de Profesores al Fin de Semestre de La Facultad de Ingeniería, Unam. Proyecto de Tesis de Grado en Ingeniería en Computación. Universidad Nacional Autónoma de México. Enero 2011.
- [9] E. González, Z. Pérez, I. Espinosa, S. Alvarezl, Obtención de Patrones y Reglas en el Proceso Académico de La Universidad de Las Ciencias Informáticas Utilizando Técnicas de Minería de Datos. Universidad de las Ciencias Informáticas.
- [10] S. Valero, Aplicación de Técnicas de Minería de Datos para Predecir Deserción. Universidad Tecnológica de Izúcar de Matamoros. 2009. [En Línea].
- [11] Krall César, Minería de datos (data mining). Qué es y para qué sirve. 2013.
- [12] S. R. Garner, “Weka: The waikato environment for knowledge analysis,” in Proceedings of the New Zealand computer science research students conference, 1995, pp.57–64.
- [13] I. Mierswa, M. Wurst., M. Klinkenberg, YALE: Rapid Prototyping for Complex Data Mining Tasks, in Proc. 12th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining (KDD-06) (2006) 935–940.
- [14] J. Gallardo, Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM. 2009. [En Línea] <<http://goo.gl/4JCF1a>>. [Acceso: 01-Diciembre-2013]
- [15] Chapman, Pete, et al. CRISP-DM 1.0 Step-by-step data mining guide. 2000.
- [16] J. Orallo, C. Ramírez, Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software.
- [17] P. Britos, V. García. Propuesta de Procesos de Explotación de Información. En XV Congreso Argentino de Ciencias de la Computación. 2009.

## **ANEXO 11: Análisis Encuesta de Evaluación de Desempeño Docente**

En el presente escrito se realizó un análisis del Cuestionario de Evaluación Docente, que se realiza en todas las áreas y carreras de la Universidad Nacional de Loja, mediante la cual, se obtienen datos estadísticos sobre la calidad del proceso enseñanza-aprendizaje, que lleva un docente dentro de su espacio de cátedra.

Partiendo de la definición dada por la Real Academia de la Lengua Española, en cuanto a cuestionario se define lo siguiente: "Lista de preguntas que se proponen con cualquier fin". De la misma fuente, se establece que evaluación es sinónimo de valoración, que define a la acción y efecto de evaluar. Esto evidencia que la evaluación especifica la estimación de un valor, en este caso a la labor docente y en el caso del cuestionario sería el acopio de datos de opinión sobre el profesor.

Cabe desatacar que el presente análisis se basó en una investigación realizada en la Universidad Nacional Autónoma de México donde previamente se llevó a cabo el análisis de del instrumento de evaluación docente.

**p:** relación del instrumento utilizado actualmente para la evaluación docente dentro de la universidad, con instrumento analizado en la investigación UNAM.

A continuación se analiza las preguntas, considerando que las mismas tienen por objetivo que el alumno haga una evaluación del desempeño de las actividades docentes de su profesor. Para ello se revisa si la posible respuesta que dé el alumno a cada una de las preguntas aporta información que efectivamente evalúe al docente como tal, tomando en consideración y como se ha dicho anteriormente que la respuesta no es abierta sino de opción múltiple (cuatro opciones).

### **Análisis de las preguntas del Instrumento:**

***1. El docente demuestra dominio científico de los contenidos que contribuyen a explicar la problemática del módulo (trata los contenidos en toda su complejidad, los relaciona, aclara y ejemplifica).***

Si es procedente la pregunta, ya que el alumno tiene la capacidad de valorar su propia perspectiva acerca de la materia (p4).

**2. El docente trabaja con contenidos pertinentes o actualizados (relacionados con los más recientes avances del conocimiento científico técnico de la profesión).**

No es procedente la pregunta, el alumno no tiene las competencias para valorar los contenidos presentados por el docente en clases. Esta evaluación es competencia de una autoridad académica (p5).

**3. Los contenidos tratados satisfacen las expectativas estudiantiles.**

Si es procedente la pregunta, ya que el alumno tiene la capacidad de valorar su propia perspectiva acerca de la materia y los resultados obtenidos al finalizarla (p4).

**4. Los contenidos seleccionados por el docente corresponden al nivel de formación del estudiante (nivel profesional) y son tratados con profundidad.**

No es procedente la pregunta, el alumno no tiene las competencias para valorar si los contenidos abarcados dentro de un módulo aportarán a su nivel profesional, ya que los mismos se encuentran aún en formación académica, pero si puede valorar si los contenidos han sido tratados con profundidad o son entendibles. (p9).

**5. El docente señala bibliografía básica y complementaria pertinente a los objetivos de la asignatura curso o taller.**

Si es procedente la pregunta, ya que el alumno puede percatarse si el profesor sugiere bibliografía como apoyo a lo aprendido en clases (p6).

**6. El docente orienta y apoya al estudiante en actividades de investigación formativa (lectura comprensiva, manejo de fuentes de consulta, identificación y planteamiento de problemas, manejo y procesamiento de datos, elaboración de conclusiones)**

Si es procedente la pregunta, ya que el alumno puede valorar la guía, apoyo e incentivo del docente hacia la investigación formativa-académica. (p6).

**7. El docente se comunica fácilmente con sus estudiantes (dialoga, presta atención, escucha, ayuda a que los estudiantes se sientan valorados, etc.)...**

Si es procedente la pregunta, ya que el alumno puede percibir la capacidad didáctica y pedagógica de su profesor para lograr una exposición clara y comprensible (p4).

**8. El docente al inicio del tratamiento del sílabo socializa con los estudiantes la programación respectiva y establece las responsabilidades de cada uno (encuadre).**

No es procedente la pregunta para las opciones enmarcadas dentro del cuestionario, ya que la misma tendría que proponer las opciones Si y No..

**9. El docente lleva a cabo un proceso lógico, ordenado y claro en el trabajo de aula y en otros escenarios.**

Si es procedente la pregunta, ya que el alumno puede apreciar la capacidad didáctica y pedagógica de su profesor para lograr una exposición clara y comprensible (p4).

**10. El docente brinda asesoría y apoyo para fortalecer los aprendizajes (organiza el trabajo en los grupos, orienta las actividades a desarrollar, amplía la información, señala las bases de datos a las que tiene acceso la Universidad).**

Si es procedente la pregunta, ya que el alumno puede percibir los apoyos didácticos a los que recurre el profesor para mejorar su aprendizaje (p18).

**11. El docente revisa periódicamente el trabajo individual y grupal que realizan los estudiantes dentro y fuera del aula.**

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo (p15).

**12. El docente asegura a través de procesos permanentes de evaluación que los estudiantes logren los aprendizajes previstos en el sílabo y utiliza diversos instrumentos de evaluación (pruebas orales, escritas, trabajos prácticos, tareas de resolución individual, de construcción grupal).**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente, con el fin de asegurar la comprensión de los contenidos abordados (p12).

**13. El docente estimula la participación activa de los estudiantes y controla sean honestos en la realización de las tareas.**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente, y a su vez observar el comportamiento ético del docente. (p8).

**14. El docente manifiesta entusiasmo y buen humor en el desarrollo de la clase**

No es procedente la pregunta, si se desea evaluar la actitud del profesor, el alumno no cuenta con las competencias para valorar los conocimientos psicopedagógicos que posee su profesor. Esta evaluación es competencia de un especialista en el tema (p3).

**15 El docente revisa los avances de los trabajos y los devuelve oportunamente con observaciones y recomendaciones.**

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo (p12).

**16. El docente lleva un registro de las actividades académicas cumplidas por los estudiantes y les informa periódicamente sobre su situación (cumplimiento de tareas, consultas, exposiciones, reportes, lecciones, informes).**

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo (p12).

**17. El docente comunica y analiza con los estudiantes los resultados de la evaluación de los aprendizajes y da la retroalimentación correspondiente.**

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo (p12).

**18. El docente planifica, supervisa y evalúa las prácticas académicas de los estudiantes (prácticas necesarias para desarrollar las habilidades y destrezas que demanda el desempeño de las prácticas profesionales).**

No es procedente la pregunta, ya que el estudiante no puede controlar las planificaciones laborales del docente, y porque además la supervisión y evaluación de las prácticas académicas está contemplado en la pregunta 15, 16, 17.

**19. El nivel de exigencia del docente en el cumplimiento de tareas es el adecuado (ajustado a las necesidades de formación profesional de los estudiantes).**

No es procedente la pregunta, el alumno no tiene los conocimientos sobre la asignatura para valorar el grado de exigencia que el docente debe tener ante las tareas expuestas, por ejemplo, el docente puede percibir que el tiempo de ejecución de un trabajo es el adecuado, más para determinado o determinados estudiantes puede ser muy poco, debido al desconocimiento de la misma (p13).

**20. El docente propicia siempre un ambiente adecuado de trabajo en el aula para que los estudiantes aprendan (ambiente de confianza, respeto y atención a los estudiantes).**

Si procede la pregunta, ya que el alumno es capaz de valorar el ambiente grupal que genera el profesor (p8).

**21. El docente motiva y hace su trabajo dinámico e interesante.**

No es procedente la pregunta, ya que no está bien dirigida, pues esto se puede visualizar de diversas maneras de acuerdo a la complejidad de los contenidos de la asignatura. En

todo caso, si se desea evaluar la actitud del profesor, el alumno no cuenta con las competencias para valorar los conocimientos psicopedagógicos que posee su profesor (p3).

**22. El docente utiliza gabinetes, laboratorios, áreas experimentales, equipos, talleres u otros escenarios en donde los estudiantes puedan vincular la teoría con la práctica (de acuerdo a lo planificado en el sílabo).**

No es procedente la pregunta ya que la ausencia y predisposición de los laboratorios en cada carrera no depende del docente, sino del área administrativa de la universidad.

**23. El docente cumple con los contenidos programados en el sílabo.**

Si es procedente la pregunta ya que el estudiante puede, observar el alcance de los contenidos programados en el sílabo.

**24. El docente es claro en sus exposiciones y explicaciones (se hace entender con facilidad).**

Si es procedente la pregunta, ya que el alumno puede percibir la capacidad didáctica de su profesor para lograr una exposición clara (p4).

**25. El docente resuelve las dudas y problemas planteados por el estudiante.**

Si es procedente la pregunta, ya que el alumno puede verificar la resolución a sus preguntas e inquietudes dentro del aula (p12).

**26. El docente logra aprendizajes significativos en los estudiantes (aprendizajes útiles para resolver problemas relacionados con las prácticas profesionales).**

No es procedente la pregunta, el alumno no tiene las competencias para valorar si los contenidos abarcados dentro de un módulo aportarán a su nivel profesional, la pregunta está enmarcada hacia lo futuro. (p9).

**27. El docente articula acciones investigativas en el desarrollo de la asignatura (consultas, procesos de análisis, de síntesis, observaciones, elaboración de informes y conclusiones).**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente. (p12).

**28. El docente incorpora las nuevas tecnologías de la información y la comunicación en el proceso enseñanza aprendizaje, para mejorar los resultados de aprendizaje del estudiante (se comunica con el estudiante a través de chats, blogs, wikis, pag. web, plataforma virtual, e-mail, video).**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente. (p12).

**29. El docente utiliza la técnica del trabajo autónomo (TAE) para la realización de algunas actividades planificadas en el sílabo y especifica con claridad las formas de interacción con los estudiantes para darles el acompañamiento adecuado.**

Si es procedente la pregunta, ya que el alumno puede percatarse si el profesor propicia que él desarrolle sus propias habilidades. (p6).

**30. El docente elabora documentos de apoyo para ampliar el conocimiento de los contenidos de la asignatura, curso o taller y los entrega a los estudiantes.**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente. (p12).

**31. La relación entre el docente y el alumno es cordial.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**32. El docente promueve en el aula y fuera de ella espacios de discusión, análisis y debate sobre temáticas relacionadas con la problemática de la profesión que se aborda en la asignatura, curso, o taller.**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente. (p12).

**33. El docente propicia aprendizajes en los estudiantes, a través de actividades de investigación formativa (los estudiantes aprenden a identificar, definir y plantear problemas; a buscar, manejar y procesar la información; a elaborar análisis y síntesis; a elaborar instrumentos para investigar; a elaborar conclusiones, informes y a trabajar en equipo).**

Si es procedente la pregunta, ya que el alumno puede apreciar la metodología de enseñanza-aprendizaje que posee el docente. (p12).

**34. El docente es ejemplo de honestidad, tolerancia y de respeto.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**35. El docente tiene un trato igual para con todos los estudiantes.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).



**36. El docente demuestra responsabilidad en su trabajo.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**37. El docente demuestra una actitud solidaria con los estudiantes ante situaciones de calamidad y de logros alcanzados.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**38. El docente demuestra lealtad y compromiso con la institución.**

No es procedente la pregunta el estudiante no puede emitir un juicio valorativo acerca de la responsabilidad del docente con la institución. Esta pregunta es competencia de una autoridad académica (p14).

**39. El docente propicia un ambiente favorable para que los estudiantes conozcan sus derechos y responsabilidades y aprendan a ejercerlos.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**40. Su trabajo en el aula contribuye efectivamente a la formación integral del estudiante.**

No es procedente la pregunta, el alumno no tiene las competencias para valorar si los contenidos abarcados dentro de un módulo aportarán a su nivel profesional. Esta pregunta es competencia de una autoridad académica (p9).

**41. El docente evalúa, acredita y califica con justicia, equidad y transparencia.**

Si es procedente la pregunta, ya que el alumno puede observar el comportamiento ético del docente (p8).

**42. El docente asiste normalmente a su trabajo.**

No es procedente la pregunta, el control de asistencia del profesor es competencia de una autoridad administrativa. El alumno no tiene la responsabilidad de verificar si el profesor cumple o no con sus obligaciones (p16).

**43. El docente cumple con responsabilidad las horas y calendarios de trabajo establecidos institucionalmente.**

No es procedente la pregunta, el control del cumplimiento del horario del profesor es competencia de una autoridad administrativa. El alumno no tiene la responsabilidad de verificar si el profesor cumple o no con sus obligaciones (p14).

***44. El docente cumple con el 100% de lo planificado.***

Si es procedente la pregunta ya que el estudiante puede, observar el alcance de los contenidos programados en el sílabo.

## ANEXO 12: Certificación Traducción Summary

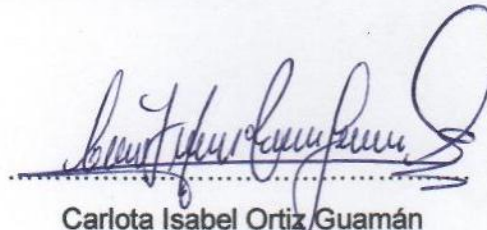
Loja, Miércoles 07 de julio del 2014

Ciudad.

Yo, Carlota Isabel Ortiz Guamán, con cedula Nro. 1102332911, Licenciada en Ciencias de la Educación con estudios en el Idioma Inglés de la Universidad Técnica Particular de Loja, respaldo que el resumen del Trabajo de Titulación traducido al idioma inglés es fiel traducción, por lo que su contenido puede ser interpretado de forma correcta.

Por su atención le expreso mi agradecimiento.

Atentamente.



Carlota Isabel Ortiz Guamán



La República del Ecuador



Y EN SU NOMBRE Y POR AUTORIDAD DE LA LEY

# La Universidad Técnica Particular de Loja

a través de sus Facultades de Ciencias de la Educación y de Lenguas y Lingüísticas

Por cuanto: **Carlota Isabel Ortiz Guamán**

Ha cumplido con los requisitos que prescribe la Ley, el Estatuto y Reglamentos Universitarios previos al Grado de Licenciado en Ciencias de la Educación y en virtud de la aprobación que obtuvo y de la promesa que prestó, le confiere el título de:

## LICENCIADO EN CIENCIAS DE LA EDUCACION ESPECIALIDAD EN INGLES

que lo habilita para ejercer las funciones inherentes a este grado académico.

Dado en Loja, a 5 de Julio de 1989

Ing. MAX TORRES GUZMAN  
RECTOR

Lic. JIMENA GUAYCHA  
CANCILLER DE LA FACULTAD DE  
CIENCIAS DE LA EDUCACION

Lic. ROSARIO BURNEO BURNEO  
DECANO DE LA FACULTAD DE  
LENGUAS Y LINGÜÍSTICAS

Dr. CARLOS RAMÍREZ ROMERO  
SECRETARIO GENERAL

CANCILLER DE LA UNIVERSIDAD  
Retirado

Acta N.º 21, Folio N.º 28  
Loja, Julio 14 de 1989

Hno. TICIANO CANCHAL GARCIA  
CANCILLER



Janina Silvana Ortiz Pasaca (2014)

Esta obra está bajo una licencia Reconocimiento-Compartir

bajo la misma licencia 3.0 de Creative Commons.

Para ver una copia de esta licencia, visite:

<http://creativecommons.org/licenses/by-sa/3.0/>