



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

DIVISIÓN DE INGENIERÍA ELÉCTRICA

TESIS PARA OBTENER EL TÍTULO DE INGENIERO EN COMPUTACIÓN

**MINERÍA DE DATOS EN ENCUESTAS DE PROFESORES AL FIN DE
SEMESTRE DE LA FACULTAD DE INGENIERÍA, UNAM**

PRESENTA

ROMÁN GARCÍA AYESHA SAGRARIO

DIRECTORES DE TESIS

Ing. GABRIELA BETZABÉ LIZÁRRAGA RAMÍREZ

Ing. JOSÉ ENRIQUE LARIOS CANALE





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mi querida alma mater, la Universidad Nacional Autónoma de México, porque gracias a ti soy lo que soy, y no me cansaré de retribuirte con mi esfuerzo, conocimiento y dedicación lo mucho que me has dado.

A Alfredo, mi esposo, por todo su apoyo durante mi carrera y por ser el motor de mi vida. Te amo.

A mis papás y hermanos, por todo su apoyo y cariño, los quiero mucho.

A mis directores de tesis y sinodales, por su tiempo y comentarios dedicados a esta tesis.

A Betzabé por todo su tiempo, paciencia y dedicación durante el desarrollo de esta tesis.

A mis compañeras de tocho, porque he aprendido mucho de ustedes; son y serán siempre, una parte importante de mi vida.

Al departamento de Control de la Facultad de Ingeniería, por confiar en mi y darme la oportunidad de trabajar con ustedes. Gracias.

Ayesha Sagrario Román García

ÍNDICE

OBJETIVO	1
INTRODUCCIÓN	2

CAPÍTULO I

INTRODUCCIÓN A LA MINERÍA DE DATOS

1.1 ¿Qué es la Minería de Datos?.....	4
1.1.1 Definición de Minería de Datos.....	4
1.1.2 Tecnología para la Minería de Datos.....	4
1.1.3 Orígenes y Problemas de la Minería de Datos.....	5
1.2 Proceso de Extracción de Conocimientos.....	5
1.2.1 Etapas del Proceso de Extracción de Conocimientos.....	5
1.2.1.1 Selección de Datos.....	6
1.2.1.2 Minería de Datos.....	7
1.2.1.3 Técnicas de Evaluación.....	8
1.2.1.4 Difusión y Uso del Conocimiento Extraído.....	8
1.2.2 Arquitectura del Proceso de Extracción de Conocimientos.....	8
1.3 Proceso de Minería de Datos.....	9
1.3.1 Fases de la Minería de Datos.....	9
1.3.2 Tareas y Herramientas de la Minería de Datos.....	9
1.4 Métodos de la Minería.....	10

CAPÍTULO II

LOS DATOS

2.1 Tipos de Datos.....	11
2.1.1 Atributos y Cantidades.....	12
2.2 Preparación de los Datos.....	12
2.2.1 Agregación.....	12
2.2.2 Muestreo.....	13
2.2.3 Reducción de Dimensionalidad.....	13
2.2.4 Función de Selección de Subconjunto.....	14

2.2.5	Discretización y Binarización.....	14
-------	------------------------------------	----

CAPÍTULO III

DESARROLLO DE LA MINERÍA DE DATOS

3.1	Tareas de Minería de Datos.....	15
3.1.1	Clasificación.....	15
3.1.1.1	Árboles de Decisión.....	16
3.1.1.2	Clasificación Basada en Reglas.....	17
3.1.1.3	Vecino más cercano.....	18
3.1.1.4	Clasificadores Bayesianos.....	18
3.1.1.4.1	Teorema de Bayes.....	18
3.1.1.4.2	Redes Bayesianas.....	19
3.1.1.5	Redes Neuronales.....	22
3.1.2	Regresión.....	24
3.1.3	Reglas de Asociación.....	25
3.1.3.1	Generación de Reglas.....	25
3.1.4	Agrupación.....	26
3.1.4.1	Tipos de Agrupación.....	26
3.1.4.2	Agrupación aglomerativa jerárquica.....	26
3.1.5	Correlación.....	28
3.2	Algoritmos de Minería de Datos.....	28
3.2.1	Algoritmos Supervisados.....	28
3.2.2	Algoritmos no Supervisados.....	29

CAPÍTULO IV

SOFTWARE LIBRE PARA LA IMPLEMENTACIÓN DE MINERÍA DE DATOS

4.1	WEKA (Waikato Environment for Knowledge Analysis).....	30
4.1.1	Historia.....	30
4.1.2	Descripción.....	30
4.1.2.1	Interfaz de Usuario.....	31
4.1.2.1.1	Explorer.....	32
4.1.2.1.2	Experimenter.....	36

4.1.2.1.3 Knowledge Flow.....	37
4.1.3 Características de WEKA.....	38
4.2 Rapid Miner.....	38
4.2.1 Historia.....	38
4.2.2 Descripción.....	38
4.2.2.1 Usuario.....	39
4.2.3 Características de Rapid Miner.....	39

CAPÍTULO V

REVISIÓN DE LA ENCUESTA DE PROFESORES AL FINAL DEL SEMESTRE EN LA FACULTAD DE INGENIERÍA, REALIZADA POR EL ING. JOSÉ ENRIQUE LARIOS CANALE

5.1 Introducción.....	40
5.2 El contexto mundial y la Educación superior.....	41
5.3 Planteamiento del Problema.....	42
5.4 Marco Jurídico Universitario.....	43
5.5 La evaluación del docente en la FI.....	44
5.5.1 Antecedentes.....	44
5.5.2 Análisis del Instrumento.....	45
5.5.2.1 ¿Evaluación o Encuesta?.....	45
5.5.2.2 Análisis de las Preguntas del Instrumento.....	47

CAPÍTULO VI

ANTECEDENTES DE MINERÍA DE DATOS

6.1 Estructura de la Encuesta de Profesores.....	51
--	----

CAPÍTULO VII

DESARROLLO DE MINERÍA DE DATOS PARA LA BASE DE DATOS DE LAS ENCUESTAS DE PROFESORES DE LA FACULTAD DE INGENIERÍA

7.1 Documentación de la Base de Datos de Encuestas de Profesores.....	55
7.1.1 Diagrama Físico Relacional.....	57
7.1.2 Diccionario de Datos Corporativo.....	57

7.2 Desarrollo de Minería de Datos.....	60
7.2.1 Selección de Datos.....	60
7.2.2 Minería de Datos.....	62
7.2.2.1 Comportamiento de los Datos.....	62
7.2.2.2 Comportamiento de los Profesores.....	63
7.2.2.3 Matriz de Correlaciones.....	66
7.2.2.4 Árboles de Decisión.....	71
7.2.2.4.1 Árboles de Decisión por División.....	78
7.2.2.4.1.1 División de Ingeniería Eléctrica.....	83
7.2.2.4.1.2 División de Ingeniería Civil.....	86
7.2.2.4.1.3 División de Ciencias Básicas.....	87
7.2.2.4.1.4 División de Ciencias Sociales y Humanidades.....	89
7.2.2.4.1.5 División de Ingeniería Mecánica e Industrial.....	91
7.2.2.4.1.6 División de Ciencias de la Tierra.....	92
7.2.2.4.2 Análisis por departamento.....	94
7.2.2.5 Agrupamiento o Clustering.....	99
7.2.2.6 Redes Neuronales.....	101
7.3 Manual del Usuario.....	105
7.3.1 Descargar RapidMiner.....	105
7.3.2 Iniciar RapidMiner.....	106
7.3.3 Trabajando con RapidMiner.....	106
CONCLUSIONES.....	110
BIBLIOGRAFÍA.....	112

OBJETIVO

Hacer uso de las diversas técnicas de la Minería de Datos, con el fin principal de extraer información útil de la Base de Datos de las encuestas de profesores, que se realizan semestre con semestre a los alumnos de la Facultad de Ingeniería, sobre las asignaturas que se encuentran cursando. La información que se utilizará corresponde a las encuestas aplicadas en el periodo comprendido del semestre 1998-1 al semestre 2010-1 y fue proporcionada por la Secretaría de Servicios Académicos de la Facultad de Ingeniería. Dicha información, será 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.

El objetivo final de la Minería de Datos es brindar las herramientas necesarias de soporte a la decisión de los directivos, en este trabajo se pretende desarrollar un DSS (Sistema de Soporte a la Decisión) dirigido principalmente a los jefes de división y de departamento que conforman a la Facultad de Ingeniería.

INTRODUCCIÓN

La Minería de Datos es un término relativamente moderno que integra numerosas técnicas de análisis de datos y extracción de modelos; todo esto con el fin de ser capaces de extraer patrones, de describir tendencias y regularidades, de predecir conceptos, y en general, de sacar provecho a la información que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades. Se define como el proceso de extracción de conocimiento válido, útil, comprensible y desconocido a partir de datos almacenados [4].

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 y corporaciones cuentan con una gran cantidad de datos históricos que detallan fielmente, la vida de esa empresa o corporación 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 la empresa.

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 empresas 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.

Pese a la popularidad del término Minería de Datos, éste es solo una etapa del proceso de extracción de conocimiento que se compone de varias fases e incorpora diferentes técnicas de los campos del aprendizaje automático, la estadística, las bases de datos, los sistemas de toma de decisión, la inteligencia artificial y otras áreas de la computación y de la gestión de la información; para ello cuenta con dos grandes rubros de tareas, descriptivas y predictivas. Las tareas descriptivas identifican patrones en los datos que los explican o resumen, mientras que las tareas predictivas estiman valores futuros de las variables de interés, a partir de otras variables independientes. Dentro de las tareas descriptivas se puede identificar al agrupamiento, asociación y correlación y dentro de las tareas predictivas se encuentran la regresión y la clasificación. Así mismo, las tareas de Minería de Datos cuentan con métodos para solucionar un determinado problema y cada método cuenta con diversos algoritmos o técnicas. En la figura a se observan las tareas de minería de datos y sus métodos.

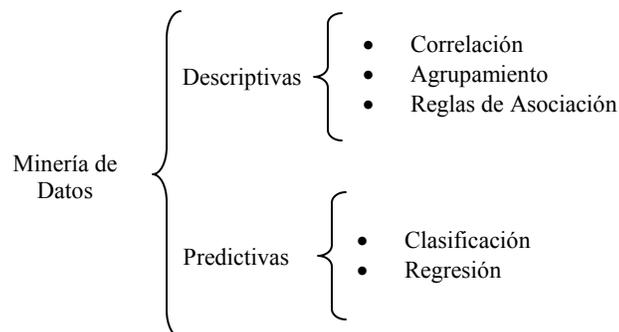


Figura a. Tareas de ,Minería de Datos

Debido al surgimiento de diversas herramientas de Minería de Datos como WEKA, RapidMiner y Clementine de SPSS que son de manejo sencillo para el usuario, se ha facilitado la popularización del uso de técnicas de Minería de Datos para no especialistas; sin embargo para profundizar sobre estas herramientas es necesario conocer más que sólo el manual de usuario, por lo que se vuelve necesario contar con un especialista que pueda sacar provecho y obtener un mejor desempeño de dichas herramientas.

Antes de comenzar con el análisis de la información y debido a la gran cantidad de datos proporcionados para esta tesis, surge la necesidad de realizar un análisis previo de los datos que consiste en conocer información adicional de los datos, por ejemplo su tamaño, tipo, etc. Posteriormente se hace un análisis previo a la minería de datos sobre la encuesta de profesores de la Facultad de Ingeniería, en el que se identifican los elementos en que se compone dicha encuesta, así como su estructura y los conceptos de evaluación que maneja sobre el profesor.

En esta tesis 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, 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 utilizará el software especializado WEKA y RapidMiner, así como SPSS para el análisis previo de los datos.

CAPÍTULO I

INTRODUCCIÓN A LA MINERÍA DE DATOS

1.1 ¿Qué es la Minería de Datos?

1.1.1 Definición de Minería de Datos

La Minería de Datos es un conjunto de técnicas y tecnologías que permiten realizar la exploración de grandes bases de datos, con la finalidad de encontrar patrones repetitivos y comprensibles, que nos permitan entender el comportamiento de los datos dentro de un contexto determinado; además reúne diferentes disciplinas como la Estadística, la Inteligencia Artificial y las Bases de Datos, haciendo uso de las herramientas y ventajas brindadas por cada una de ellas, con el fin de descubrir conocimiento novedoso y potencialmente útil dentro los datos[1].

El objetivo final de la Minería de Datos es brindar las herramientas necesarias de soporte a la decisión de los directivos empresariales, en este trabajo se pretenden desarrollar las herramientas dirigidas principalmente a los jefes de división y de departamento que conforman a la Facultad de Ingeniería.

1.1.2 Tecnología para la Minería de Datos.

Existe una amplia variedad en el mercado de tecnología para la minería de datos, y se encuentran clasificadas en tres grandes categorías:

1. Análisis estadístico o de datos.
Se usan para detectar patrones no usuales en los datos, y estos se explican mediante modelos estadísticos y matemáticos. Algunas de las técnicas de modelado estadístico son: análisis lineal y no lineal, análisis de regresión continua y logística, entre otros.
2. Proceso de Extracción del Conocimiento (KDD).
Se utiliza cuando los datos son muy extensos y tienen una gran cantidad de variables, tales que vuelven complejo el análisis estadístico de los datos; tiene su origen en la inteligencia artificial y el aprendizaje con máquinas.
3. Sistemas de visualización, sistemas de información geográfica y análisis fractal.

1.1.3 Orígenes y Problemas de la Minería de Datos.

La Minería de Datos surge con el aumento de volumen y variedad de los datos, ya que utiliza información que puede ser histórica, de transacciones o situaciones que se han producido con la intención de predecir la información futura.

Muchos investigadores de diferentes disciplinas han desarrollado e implementado sus herramientas para el manejo de diversos tipos de datos; este trabajo ha culminado en el desarrollo de herramientas para la Minería de Datos, construyendo una metodología y algoritmos que los investigadores habían usado previamente [4].

La Minería de Datos ha adoptado ideas de otras disciplinas como la optimización, computación evolutiva, teorías de la información, procesamiento de señales y visualización, con el fin de contar con una amplia gama de algoritmos para resolver los problemas que presenta esta novedosa disciplina [7].

1.2 Proceso de Extracción del Conocimiento.

La Minería de Datos forma parte del proceso de Descubrimiento de Conocimientos (KDD), que consiste en hacer uso de algoritmos que nos permiten extraer de una base de datos, lo que llamamos Conocimiento (datos ocultos, previamente desconocidos y potencialmente útiles), mediante la especificación de parámetros adecuados; el proceso de Extracción de Conocimiento se lleva a cabo de forma interactiva y repetitiva, esto se refiere a que el usuario debe tener participación en todo el proceso, ya que es él mismo el que da el contexto a los datos y determina que parte de los datos desea utilizar [2].

1.2.1 Etapas del Proceso de Extracción de Conocimiento.

El proceso de Extracción de Conocimiento está compuesto de los siguientes pasos:

- Selección de Datos.
- Minería de Datos.
- Técnicas de evaluación y mejora de modelos.
- Difusión y uso del conocimiento extraído.

En la figura 1.1 se muestran las etapas del Proceso de Extracción de Conocimiento (KDD):



Figura 1.1 Proceso de Extracción de Conocimiento

1.2.1.1 Selección de Datos.

La selección de datos relevantes es uno de los preprocesamientos más importantes, ya que es crucial que los datos utilizados sean relevantes y útiles para la tarea de Minería de Datos abordada la cual puede ser horizontal o vertical¹. Cuando la selección de datos es horizontal, sólo se eligen instancias completas representativas del total de los datos disponibles.

Algunos ejemplos de estos tipos de muestreo son:

- a) Aleatorio Simple (con o sin reemplazamiento). En este tipo de muestreo todos los elementos en un conjunto de datos, tienen la misma probabilidad de ser seleccionados. La selección de muestras puede realizarse a través de cualquier mecanismo probabilístico.
- b) Aleatorio Estratificado. En este tipo de muestreo se divide a los datos en una serie de grupos llamados estratos y de cada estrato se toma una muestra aleatoria. El criterio para realizar la división de estratos consiste en identificar grupos homogéneos respecto a la característica a estudiar.
- c) De Grupos o Conglomerados. En este tipo de muestreo, se divide a los datos en conjuntos, de tal forma que cada conjunto tenga características en común; una vez que los conjuntos o grupos han sido seleccionados, se toma una muestra aleatoria de cada grupo.
- d) Exhaustivo. En el muestreo exhaustivo se subdividen los datos en varios niveles ordenados que se extraen sucesivamente por medio de un procedimiento de embudo. El muestreo se desarrolla en varias fases o extracciones sucesivas para cada nivel; un ejemplo de este tipo de muestreo se presenta cuando se tiene un conjunto de prácticas médicas y se seleccionan muestras aleatorias sin regresarse al conjunto de prácticas, de esta forma irán surgiendo categorías según el tema del que se trate la práctica; el muestreo termina cuando la información obtenida de las prácticas o muestras resulta redundante [3].

Cuando la selección de datos es vertical la idea es seleccionar los atributos más relevantes de las instancias en base a algún criterio general o dependiente del problema particular. Como criterios generales para la selección de datos vertical, podemos citar la eliminación de distintas formas de llaves primarias (códigos de identificación internos, nombres y apellidos, códigos postales) y la eliminación de atributos dependientes (ciudades, direcciones, sexo y teléfonos).

La jerarquía del conocimiento o pirámide del conocimiento, es un conjunto de modelos que ayudan a entender la relación que existe entre datos, información y conocimiento; este

¹ Los atributos o campos seleccionados, deben ser significativos para el conocimiento que se desea obtener y la tarea de Minería de Datos que se pretende realizar.

concepto puede ayudar durante el análisis y la selección de datos. En la figura 1.2 se muestra la jerarquía del conocimiento.

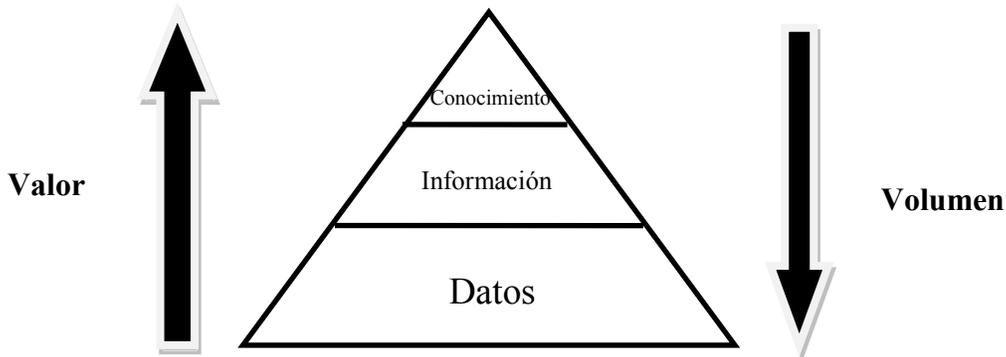


Figura 1.2 Jerarquía del Conocimiento

1.2.1.2 Minería de Datos.

El proceso de Minería de Datos consiste en el uso de tareas de Minería de Datos y en el uso de algoritmos concretos que generan patrones sobre los datos procesados. Las tareas de Minería de Datos, se dividen en 5 tipos para la resolución de problemas de negocios: clasificación, regresión, reglas de asociación, agrupación y correlación las cuales son desarrolladas durante la sección 3.1 de la presente tesis. Algunas tareas como la clasificación y regresión son principalmente utilizadas para la predicción, mientras que la agrupación puede ser usada tanto para la predicción como para la descripción [1].

El proceso de Minería de Datos se divide en 4 etapas principales como se muestra en la tabla 1.1:

Fase	Nombre	Descripción
1	Data Warehousing	Sistema de Gestión de Datos de Soporte a la decisión. Recopilación, limpieza y transformación de datos para sistemas operativos y para la creación del Almacén de Datos.
2	Minería de Datos	Reúne diferentes disciplinas como la Estadística, la Inteligencia Artificial y las Bases de Datos, con el fin de descubrir conocimiento novedoso y potencialmente útil dentro de grandes cantidades de datos.
3	Modelos Predictivos	Tarea frecuente de la Minería de Datos, la cual tiene como objetivo la generación de modelos precisos e interpretables.
4	Predicción de Estimaciones	Consiste en obtener estimaciones o pronósticos de valores futuros de un atributo.

Tabla 1.1 Proceso de Minería de Datos

1.2.1.3 Técnicas de Evaluación.

El proceso de evaluación e interpretación se realiza cuando se han adquirido mediante el proceso de Minería de Datos, los modelos o patrones de los datos y se desea evaluar los siguientes aspectos:

- Los modelos o patrones son precisos.
- El modelo extraído es comprensible.
- La información obtenida es útil y novedosa.

La evaluación de los modelos se puede realizar de las siguientes formas:

1. Validación Simple. Durante este proceso se dividen los datos en dos conjuntos, el primero será utilizado para hacer pruebas, mientras que el segundo será utilizado para la instanciación del modelo.
2. Validación Cruzada. Este se lleva a cabo cuando el conjunto de datos es muy pequeño, su proceso consiste en dividir los datos en subconjuntos pequeños y de ellos obtener el porcentaje de error de cada uno de ellos, de esta manera al sacar el promedio de los porcentajes de error de los subconjuntos, se obtiene una mejor precisión sobre la información [7].

1.2.1.4 Difusión y Uso del Conocimiento Extraído.

En esta etapa del Proceso de Extracción de Conocimientos (KDD) se pueden tomar decisiones con base en la información obtenida; los modelos obtenidos pueden ser utilizados en otros sistemas y éstos deben de ser verificados constantemente en cuanto a su funcionalidad.

1.2.2 Arquitectura del Proceso de Extracción de Conocimientos.

La Minería de Datos forma parte integral del Proceso de Extracción de Conocimientos en las bases de datos (KDD). Este proceso consiste en una serie de pasos de transformación, desde el preprocesamiento de los datos hasta el posprocesamiento de los resultados de la Minería de Datos. En la figura 1.3 se muestra la arquitectura del Proceso de Extracción de Conocimientos:

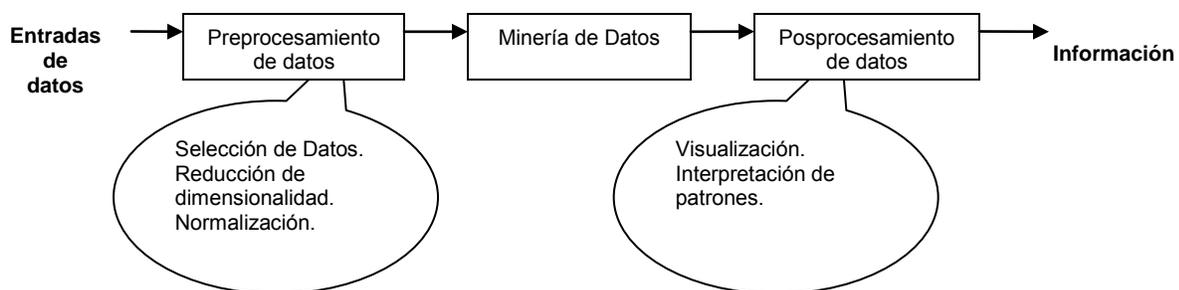


Figura 1.3 Arquitectura para la Extracción del Conocimiento

La entrada de los datos, puede ser en una variedad de formatos (archivos de texto plano, hojas de cálculo, o tablas relacionales), y pueden encontrarse en un repositorio de datos centralizado o distribuido en múltiples nodos. El propósito del preprocesamiento, es la transformación de los datos de entrada en un formato apropiado para su subsecuente análisis. Para la integración de la información obtenida se requiere de un posprocesamiento de datos, paso en el que los resultados son validados y son incorporados al Sistema de Soporte a la Decisión (DSS) [4].

1.3 Proceso de Minería de Datos.

Durante el proceso de la Minería de Datos se busca que los datos se describan así mismos e indiquen cuáles son los atributos que los caracterizan; después, se valida la hipótesis obtenida a partir de los datos.

La minería de datos es una tecnología compuesta por etapas que integra varias áreas. Durante el desarrollo del proceso de Minería de Datos se pueden utilizar diferentes aplicaciones de software para cada etapa, que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente. Actualmente existen aplicaciones o herramientas comerciales de minería de datos muy poderosas que contienen un sinnúmero de utilerías que facilitan el desarrollo de un proyecto.

En el proceso de Minería de Datos se realiza un análisis de las tareas 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.

1.3.1 Fases de la Minería de Datos.

El proceso de la Minería de Datos se encarga de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático y otras herramientas de diversas disciplinas para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

La Minería de Datos permite a los individuos comprender y modelar de una manera más eficiente el contexto en el que deben tomar decisiones, de esta forma desarrolla diferentes tipos de tareas como las asociaciones, agrupamientos, clasificaciones, análisis correlacional, entre otras. Un ejemplo de asociación se realiza cuando en un supermercado se colocan las cosas que las personas compran al mismo tiempo, por ejemplo pañales y cervezas en el que si un hombre compra pañales, también comprará cerveza, entonces se colocan las cervezas y los pañales juntos para que aumente la venta de estos productos.

1.3.2 Tareas y Herramientas de la Minería de Datos.

Las tareas de Minería de Datos son el resultado de un largo proceso de investigación y desarrollo de productos principalmente para las grandes organizaciones. Esta evolución comenzó cuando los datos de los negocios fueron almacenados por primera vez en computadoras, continuó con mejoras en el acceso a los datos, y más recientemente con

tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. La Minería de Datos toma este proceso de evolución más allá del acceso y retrospectiva de los datos, hacia la entrega de información para predecir situaciones futuras y toma de decisiones.

La Minería de Datos se encuentra soportada por tres grandes herramientas:

1. Almacén de Datos (recolección masiva de los datos).
2. Computadoras con multiprocesadores.
3. Algoritmos de Minería de Datos.

1.4 Métodos de la Minería de Datos.

Algunos de los métodos utilizados en la Minería de Datos son los siguientes:

- Árboles de Decisión. Este método realiza cortes sobre una variable lo cual, limita su expresividad pero facilita su comprensión.
- Clasificación y Regresión. Ajustan las combinaciones de funciones lineales y no lineales.
- Basados en Ejemplos Prototípicos. Se hacen aproximaciones y a los ejemplos más conocidos.
- Modelos Gráficos de Dependencias Probabilísticas. Dado un modelo estructural y probabilístico, se encuentran los valores de ciertas variables.

CAPÍTULO II LOS DATOS

2.1 Tipos de Datos

Antes de comenzar con el análisis de Minería de Datos, se hace necesaria la creación de una Base de Datos, la cual albergará toda la información obtenida sobre las encuestas de los profesores desde el semestre 1998-1 hasta el semestre 2010-1, para la cual es necesario conocer los tipos de datos que pueden ser almacenados, así como para la selección de los mismos.

En una Base de Datos existen diversos tipos de datos, los objetos que contienen datos tienen asociado un tipo de dato que define su dominio, por ejemplo, el objeto puede contener carácter, entero o binario. Los siguientes objetos tienen tipos de datos:

- Columnas de tablas y vistas.
- Parámetros de procedimientos almacenados.
- Variables.
- Procedimientos almacenados que devuelven un código, que siempre es de tipo entero.

A continuación se describe un listado de algunos tipos de datos más utilizados en lenguaje de programación PL/SQL:

1. Tipo de dato Numérico.

- NUMBER(P,S). Este tipo de dato puede contener un valor numérico de tipo entero o flotante, donde P es la cantidad de dígitos del valor numérico y S es la cantidad de dígitos después del punto decimal.

2. Tipo de dato Carácter.

- VARCHAR2(L). Este tipo de dato puede almacenar una cadena de caracteres variable de tamaño máximo L.
- CHAR(L). Este tipo de dato es similar a VARCHAR la diferencia es que su cadena de caracteres es fija y los espacios que no son utilizados son rellenados con caracteres en blanco.

3. Tipo de dato Fecha.

- DATE. Este tipo de dato almacena información sobre el día, mes, año, hora, minuto y segundo.
- TIMESTAMP (P). Cuenta con las mismas características de DATE pero además tiene la capacidad de almacenar fracciones de segundo, donde P es el número de cifras de las fracciones de segundo.

2.1.1 Atributos y Cantidades

Un atributo representa una característica de interés sobre una entidad específica y su elección depende del uso que se le dará a la Base de Datos. Los atributos de los datos se generan cuando a un objeto le es asignado un tipo de dato específico, definiendo de esta forma cuatro atributos principales:

- Tipo de dato del objeto.
- Longitud o tamaño del valor almacenado.
- Escala y/o la precisión del número (tipo de dato numérico).

2.2 Preparación de los Datos

La preparación de los datos en un proyecto de Minería de Datos es indispensable para obtener buenos resultados; consiste en realizar un preprocesamiento de los datos antes de ser utilizados para construcción de una vista minable². El objetivo principal de la preparación de los datos es realizar una estandarización para interdependencia e impedir que exista una mezcla de tipos de datos y mayor independencia entre ellos. El objetivo de la preparación de los datos es poder aplicar la técnica de Minería de Datos que más se adecúe al problema y a los datos mismos.

Para poder analizar los datos, es necesario en principio, contar con ellos; es por eso que la preparación de los datos incluye la recopilación de datos, etapa en la que, debido a la gran diversidad y tamaño de las fuentes, se requiere de la toma de decisiones en las que se elige el tipo de fuentes, internas o externas, organización, mantenimiento de datos en relación al tiempo, y finalmente de qué forma se van a poder extraer parcial o totalmente, en detalle o agregados, con distintas vistas minables a las que podamos aplicar las distintas herramientas de minería de datos [7].

2.2.1 Agregación

La agregación se define como la combinación de dos o más objetos dentro de uno solo; es considerada una herramienta de datos que se compone de una serie de transacciones llamadas objetos de datos.

Desde este punto de vista, la agregación es un proceso de eliminación de atributos o de reducir el número de valores para un atributo en particular. Por ejemplo, reducir los posibles valores de las fechas de 365 días a 12 meses, es un tipo de agregación que comúnmente se utiliza en el Procesamiento en Línea Analítico (OLAP).

² Es una única tabla que contiene todos los atributos relevantes para Minería de Datos

En la tabla 2.1 se puede observar un arreglo multidimensional, donde cada atributo es una dimensión:

TransacciónID	Artículo	Localidad	Fecha	Precio	...
.
.
101123	Reloj	Chicago	09/06/04	\$25.99	...
101124	Batería	Chicago	09/06/04	\$5.99	...
101125	Zapatos	Minneapolis	09/06/04	\$75.00	...
.
.
.

Tabla 2.1 Arreglo Multidimensional

Algunas ventajas de utilizar la agregación es que los pequeños conjuntos de datos resultantes de la reducción de datos, requieren menos memoria y menos tiempo de procesamiento. Además la agregación permite el uso de algoritmos de Minería de Datos menos costosos debido a la reducción de los datos.

2.2.2 Muestreo

El muestreo es utilizado para hacer aproximaciones de selección de una subcategoría de datos para ser analizados. En la estadística, el muestreo es utilizado durante las investigaciones preliminares de los datos y en el análisis final de los datos. Sin embargo las ventajas de utilizar el muestreo en la estadística y en la minería de datos, son diferentes. La estadística usa el muestreo para obtener la herramienta de datos de interés y saber si es costosa o consume mucho tiempo, mientras que los mineros de datos, obtienen la herramienta de datos y determinan si es costosa o consume mucho tiempo, durante el muestreo sobre los datos. Es decir que en la estadística se utiliza el muestreo antes de utilizar la herramienta deseada y determina si es costosa o no, mientras que en la minería de datos se determina si la herramienta es costosa o no después de la ejecución de ésta sobre todos los datos.

La clave para realizar un muestreo efectivo es el siguiente:

Utilizando una muestra se puede trabajar de la misma forma que utilizando todos los datos si la muestra es representativa. En cambio, una muestra es representativa si se tienen las mismas propiedades de interés como en los datos originales. Si el porcentaje de objetos de datos es la propiedad de interés, entonces una muestra es representativa si su porcentaje es cercano a los datos originales.

Debido a que el muestreo es un proceso estadístico, la representatividad de una muestra en particular puede variar, y lo mejor que podemos hacer es elegir un esquema de muestreo que garantice una amplia probabilidad de obtener una muestra representativa.

2.2.3 Reducción de Dimensionalidad

Los conjuntos de Datos pueden tener un número muy grande de atributos. Considerando un conjunto de documentos, en el que cada documento es representado por un vector cuyos

componentes son la frecuencia con la que cada palabra se presenta en el documento. En este caso, se tienen cientos de atributos o componentes, uno por cada palabra en el vocabulario.

Se tienen una gran variedad de beneficios de la reducción de Dimensionalidad³; uno de los principales beneficios es que muchos algoritmos de Minería de Datos, trabajan mejor si la Dimensionalidad (el número de atributos en los datos) es baja. Esto es en parte porque la reducción de Dimensionalidad puede eliminar características irrelevantes y reducir el ruido.

2.2.4 Función de Selección de Subconjunto

Otra forma de reducir la Dimensionalidad es utilizar solamente un subconjunto de características. Las características redundantes duplican mucha o toda la información contenida en uno o más atributos. Por ejemplo, el precio de un producto, la cantidad de ventas y el tipo de pago contienen mucha de la misma información. Las características irrelevantes contienen información que no es utilizada para la minería de datos. Por ejemplo, el número de identificación de un estudiante es irrelevante para la tarea de predecir el promedio de calificaciones del estudiante.

Cuando algunos atributos son irrelevantes o redundantes pueden ser eliminados inmediatamente por sentido común o por el dominio del tema, seleccionar el mejor subconjunto de atributos comúnmente requiere de una aproximación sistemática. La aproximación ideal para la selección de características, es probar todos los subconjuntos de atributos como entradas para los algoritmos de minería de datos de interés, y se toma el subconjunto que produce mejores resultados.

2.2.5 Discretización y Binarización

Algunos algoritmos de minería de datos, especialmente algunos algoritmos de clasificación, requieren que los datos se encuentren en forma categórica de atributos. Los algoritmos que encuentran patrones de asociación requieren que los datos se encuentren en forma binaria de atributos. De esta forma, a menudo es necesaria una transformación continua de los datos en atributos categóricos (discretización), y los atributos discretos y continuos pueden necesitar la transformación de uno o más atributos binarios (binarización).

Adicionalmente, si los atributos categóricos tienen un número grande de valores (categorías), o algunos valores ocurren esporádicamente, es más seguro o benéfico para las tareas de minería de datos, reducir el número de categorías y combinaciones posibles de valores.

³ Se refiere al número de atributos en los datos.

CAPÍTULO III DESARROLLO DE LA MINERÍA DE DATOS

3.1 Tareas de Minería de Datos

Dentro de la Minería de Datos, hemos de distinguir tipos de tareas de Minería de Datos, cada una de las cuales puede resolver problemas por medio de un algoritmo de Minería de Datos. Esto significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra.

Las distintas tareas pueden ser predictivas o descriptivas. Entre las tareas predictivas encontramos la clasificación y la regresión, mientras que en las tareas descriptivas encontramos el agrupamiento, reglas de asociación y correlaciones.

3.1.1 Clasificación

La tarea de clasificación se utiliza donde existe un problema de asignación de objetos en categorías estrictamente predefinidas, y de esta forma puede tener infinidad de aplicaciones. Ejemplos de éstas es la detección de mensajes de spam en el correo electrónico, basado en la detección de encabezados y contenidos, la clasificación de las galaxias basada en sus sombras, la clasificación de los medicamentos para saber cuál es el mejor para una enfermedad en particular, entre otras.

Las entradas de datos para un problema de clasificación es una colección de datos. Cada colección es conocida como instancia o ejemplo, es caracterizada por una tupla (x,y) donde x es un atributo auxiliar y es un atributo especial.

La tabla 3.1 muestra un conjunto de datos utilizados para la clasificación de vertebrados en una de las siguientes categorías: mamíferos, aves, peces, reptiles, o anfibios.

Nombre	Temp. Corporal	Piel	Criatura Acuática	Criatura Aérea	Tiene Piernas	Hiberna	Clase
Humano	Sangre caliente	pelo	No	No	Si	No	Mamífero
Víbora	Sangre fría	escamas	No	No	No	Si	Reptil
Salmon	Sangre fría	escamas	Si	No	No	No	Pez
Rana	Sangre fría	nada	Semi	No	Yes	Yes	Anfibio
Gato	Sangre caliente	pelo	No	No	Yes	No	Mamífero
Paloma	Sangre caliente	plumas	No	Si	Si	No	Ave
Tortuga	Sangre fría	escamas	Semi	No	Si	No	Reptil
Pinguino	Sangre caliente	plumas	Semi	No	Si	No	Ave

Tabla 3.1 Datos de algunos animales vertebrados

3.1.1.1 Árboles de Decisión

La elaboración de árboles de decisión se ha utilizado como método de Minería de Datos para representar los eventos que surgen a partir de una decisión. Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final se puede tomar siguiendo una serie de condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Los árboles de decisión han sido utilizados principalmente para explicar procedimientos médicos, legales, matemáticos, estratégicos, entre otros.

En la figura 3.1 siguiente se muestra un ejemplo de árbol de decisión para determinar la recomendación o no de cirugía ocular a un paciente; en el primer nodo se pregunta si el paciente tiene astigmatismo, en el caso de que no tenga astigmatismo se pregunta si tiene miopía y dependiendo del valor de miopía de determina si requiere cirugía o no, si el paciente tiene astigmatismo se requiere saber el valor que tiene de astigmatismo, en caso de que tenga >25 y <50 se pregunta si tiene miopía y se toma una decisión.

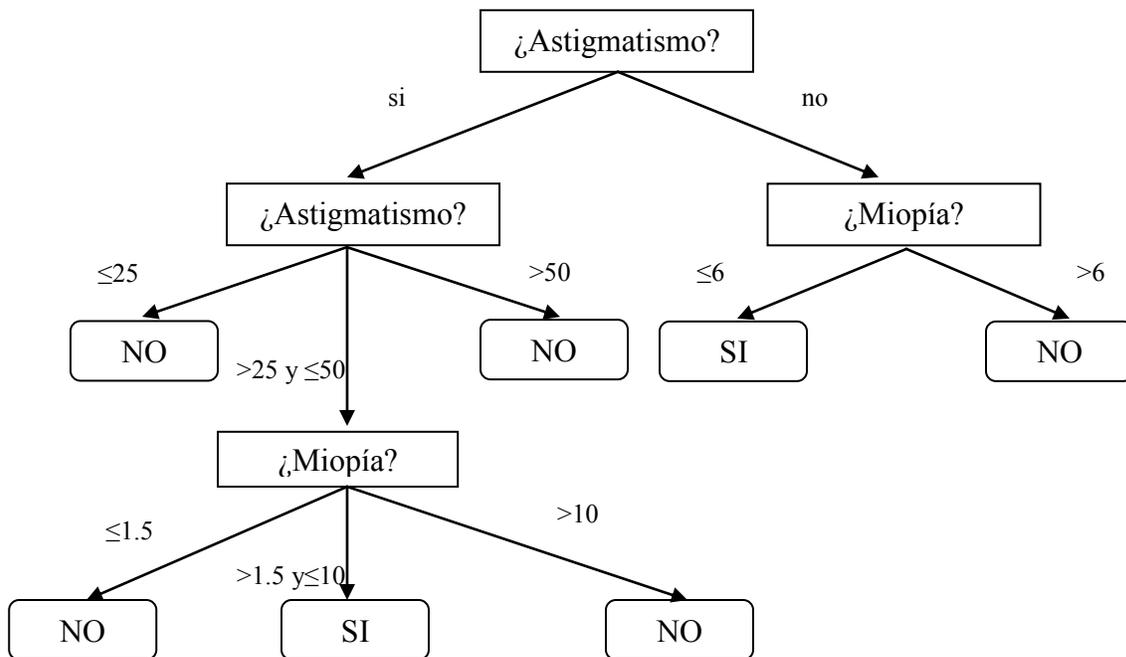


Figura 3.1 Árbol de decisión de cirugía ocular

La tarea de aprendizaje⁴ a la cual se adecuan mejor los árboles de decisión es la clasificación. La clasificación tiene como propiedad que asume que las clases son disjuntas, es decir, que una distancia es de la clase a o es de la clase b, un ejemplo de clasificación sería determinar si una luz en el firmamento es una estrella o planeta: es una cosa o es otra, y además esta propiedad es exhaustiva, es decir que se tiene cumplir una de las dos

⁴ Obtener conocimiento mediante un estudio o experiencia. Existen 3 tipos: Supervisado, No supervisado y por refuerzo.

condiciones. Dio lugar a algunos de los primeros algoritmos de árboles de decisión, a éstos se les llamó algoritmos de partición o de “divide y vencerás”. Uno de los aspectos más importantes de los árboles de decisión es el criterio, y esto ayudó a mejorar los primeros algoritmos, ya que en los primeros algoritmos si se tomaba una mala decisión, ésta no podía cambiarse, y en su mayoría, esto generaba un peor árbol⁵.

3.1.1.2 Clasificación Basada en Reglas

Las reglas expresan patrones de comportamiento entre los datos en función de la aparición conjunta de valores de dos o más atributos. Estos patrones pueden servir para conocer el comportamiento en general de la base de datos, por ejemplo en la toma de decisiones. Un ejemplo en un supermercado, en el que podemos conocer que productos suelen comprarse conjuntamente, y así mejorar la distribución de productos en estanterías; otro ejemplo sería un servidor Web, en el que podríamos conocer cuáles son las rutas de visita de un usuario al consultar una página Web y así mejorar la estructuración de las páginas Web en ese servidor.

Una regla es una proposición probabilística sobre la ocurrencia de ciertos estados en una base de datos. Las métricas de calidad de una regla, suelen ser las siguientes:

- Cobertura. Se refiere al número de instancias que la regla predice correctamente.
- Confianza. Mide el porcentaje de veces que la regla se cumple cuando se puede aplicar.

Un ejemplo de ello, son las siguientes reglas obtenidas de la tabla 3.1:

R1: (temperatura corporal = sangre fría) → No es mamífero
R2: (temperatura corporal = sangre fría)^(criatura acuática = si) → Pez
R3: (temperatura corporal = sangre caliente)^(piernas= si) → mamífero

La cobertura de la regla R1 es buena ya que al aplicarla sobre la tabla 3.1 predice correctamente todos los casos; la confianza de la regla R1 es 100% ya que se cumple al aplicarla todas las veces posibles sobre la tabla 3.1.

Para la regla R2 se obtiene un resultado igual al de la regla R1; observando la regla R3 podemos verificar que algunos vertebrados con sangre caliente y que tienen piernas, también pueden ser aves, por lo que la cobertura se reduce al 50% en comparación con la regla R1, así su confianza disminuye al 50% ya que solo en la mitad de los casos en los que puede ser aplicada se cumple la regla R3.

El aprendizaje de las reglas se divide en dos etapas: la extracción de conjunto de elementos que cumplan con la cobertura requerida y la generación de las reglas a partir de estos conjuntos.

⁵ Al tomar una mala decisión y no poder corregirla se obliga al algoritmo a tomar malas decisiones subsecuentes y como resultado nos da un árbol inútil en la toma de decisiones.

3.1.1.3 Vecino más cercano

Este algoritmo de clasificación, pertenece al aprendizaje basado en instancias o casos, las instancias se almacenan en memoria, de tal forma que cuando llega una nueva instancia cuyo valor es desconocido se intenta relacionar ésta con las instancias almacenadas (cuya clase o valor es conocida) buscando las que más se parecen. Todo el trabajo se realiza cuando llega una nueva instancia a clasificar y no cuando se procesa el conjunto de entrenamiento; de esta forma podemos decir que se trata de un método perezoso o retardado, ya que retrasa el trabajo real tanto como sea posible.

En este tipo de aprendizaje, cada instancia nueva se compara con las existentes usando una métrica de distancia, y la distancia más próxima se une para asignar su clase a la instancia nueva. Este tipo de clasificación es muy útil para trabajar con tipos de datos no estándar como texto o multimedia. El único requerimiento para incluir un tipo de dato es la existencia de una métrica apropiada para formalizar el concepto de similitud.

3.1.1.4 Clasificadores Bayesianos

El fundamento principal de estos clasificadores es la suposición de que todos los atributos son independientes, una vez conocido el valor del atributo llave. A pesar de que realiza esta suposición, el clasificador demasiado fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados y además demuestra que sus resultados son competitivos con otras técnicas en muchos problemas y que incluso las superan en muchos otros [3].

La hipótesis de independencia asumida por el clasificador genera un modelo gráfico en el que existe un único nodo raíz (la clase) y en la que todos los atributos son nodos hoja que tienen como único padre a la variable clase.

3.1.1.4.1 Teorema de Bayes

El teorema de Bayes es la regla básica para realizar inferencias, de tal forma permite actualizar la creencia que se tiene sobre un suceso o conjunto de sucesos al observar nuevos datos, es decir, permite pasar de la probabilidad a priori $P(\text{suceso})$ a la probabilidad a posteriori $P(\text{suceso}/\text{observaciones})$.

Teniendo en cuenta estos conceptos, el teorema de Bayes puede ser representado por la siguiente expresión:

$$P(h/O) = \frac{P(O/h) \cdot P(h)}{P(O)}$$

Donde, lo que aparecen son la probabilidad a priori de la hipótesis (h) y de las observaciones (O) y las probabilidades condicionadas $P(h/O)$ y $P(O/h)$. A esta última se le conoce como verosimilitud de que la hipótesis h haya producido el conjunto de observaciones O.

De esta forma, el teorema de Bayes nos permite de una forma sencilla y con semántica clara resolver las tareas de clasificación de minería de datos. Sin embargo, su desventaja es que requiere de complejidad computacional, debido a que se necesita trabajar con muchas probabilidades lo que genera muchas variables, haciéndolas, en su mayoría, inmanejables. Para resolver este problema, algunas veces se hace uso de la supuesta independencia de variables.

Un ejemplo de Teorema de Bayes es el siguiente:

El 20% de los empleados de una empresa son ingenieros y otro 20% son economistas. El 75% de los ingenieros ocupan un puesto directivo y el 50% de los economistas también, mientras que los empleados que no son ingenieros o economistas solamente el 20% ocupan un puesto directivo. ¿Cuál es la probabilidad de que un empleado directivo elegido al azar sea ingeniero?



$$p(\text{ingeniero/directivo}) = \frac{0.2 \cdot 0.75}{0.2 \cdot 0.75 + 0.2 \cdot 0.5 + 0.6 \cdot 0.2} = 0.405$$

La probabilidad de que un empleado seleccionado al azar se ingeniero es 40.5%.

3.1.1.4.2 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. Las Redes Bayesianas representan el conocimiento cualitativo del modelo mediante un grafo⁶ dirigido acíclico. Este conocimiento se articula mediante las relaciones de independencia/dependencia de los atributos o variables que componen el modelo. Un aspecto muy importante para la Minería de Datos, es el hecho de utilizar una representación gráfica para la especificación del modelo, lo que hace a las Redes Bayesianas una herramienta muy atractiva.

⁶ Es un conjunto de nodos y una selección de aristas que pueden estar orientados o no.

En la figura 3.2 se muestra el modelo gráfico de los clasificadores Bayesianos:

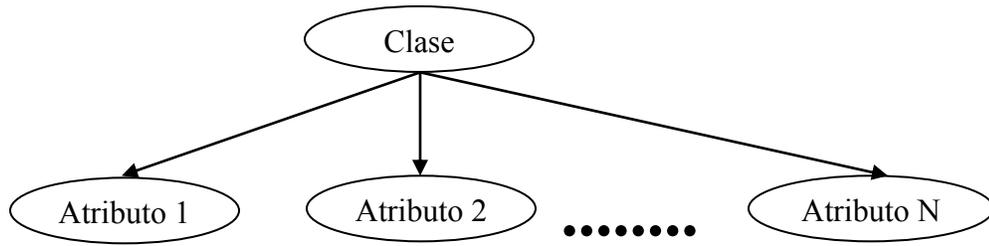


Figura 3.2 Clasificador Bayesiano

En la figura 3.3 se muestra un ejemplo de modelo gráfico de clasificador Bayesiano construida sobre una base de datos de Ingresos:

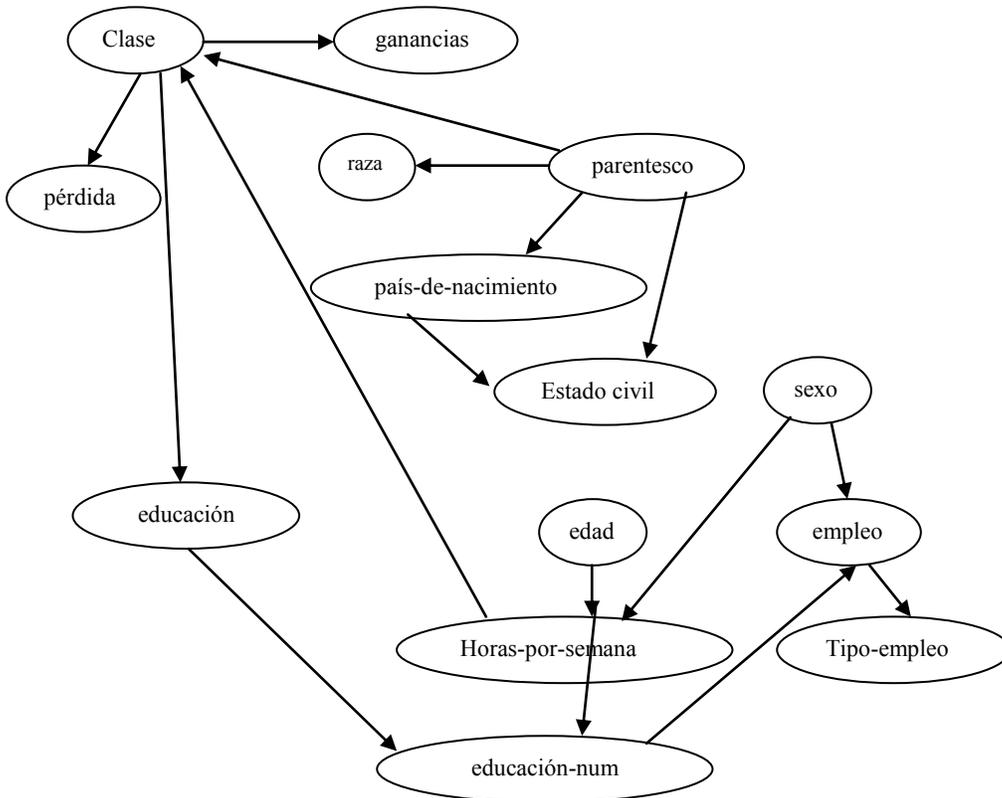


Figura 3.3 Clasificador Bayesiano de Ingresos

En la figura 3.4 se muestra un ejemplo de una Red Bayesiana en que se determina si un cliente comprará o no una computadora personal:

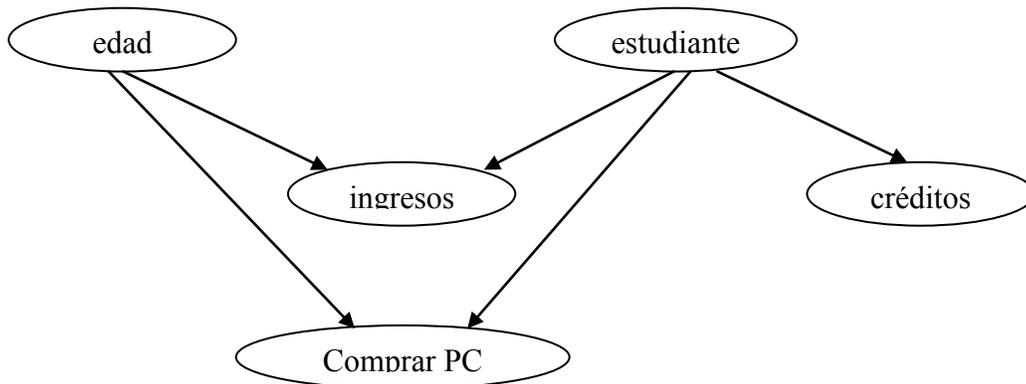


Figura 3.4 Ejemplo de Red Bayesiana

Una vez definido un modelo como el de la figura anterior, podremos realizar un proceso de inferencia. En la figura 3.5 se muestra un esquema de probabilidades a priori, así vemos que en principio el 60 por ciento de la población compraría una PC.

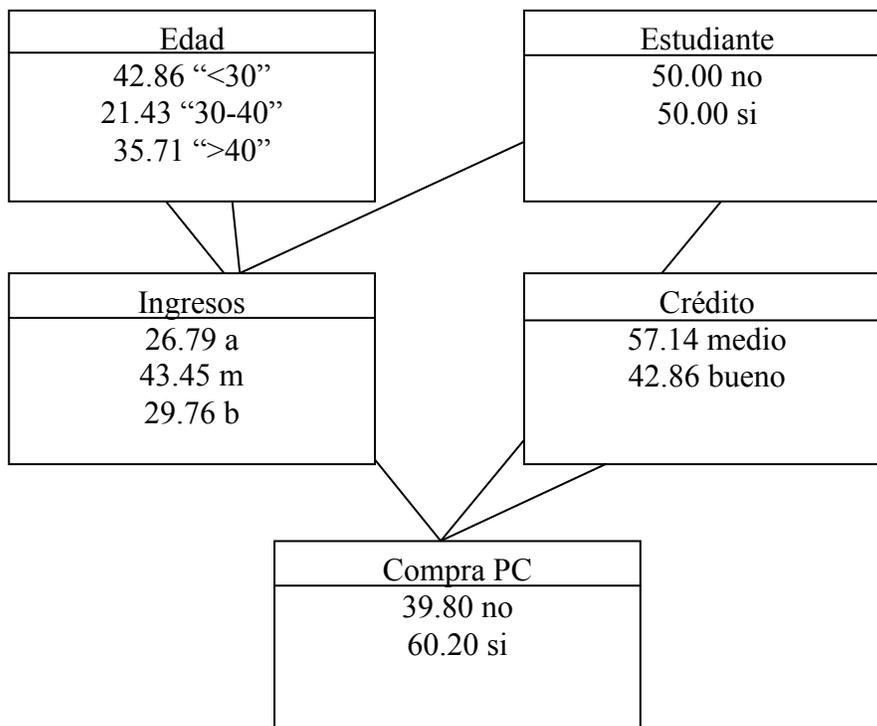


Figura 3.5 Esquema de Probabilidades A priori de compra de PC

Sin embargo, nosotros podemos conocer una evidencia, como por ejemplo que la persona en cuestión es un estudiante, entonces la probabilidad de la variable estudiante cambiaría al

100 por ciento, conocimiento totalmente cierto, y entonces el resto de las variables cambiaría en función de la observación que hemos realizado.

En este caso, la probabilidad de compra de una PC, sabiendo que la persona es un estudiante, pasa a ser de un 79 por ciento como se muestra en la figura 3.6:

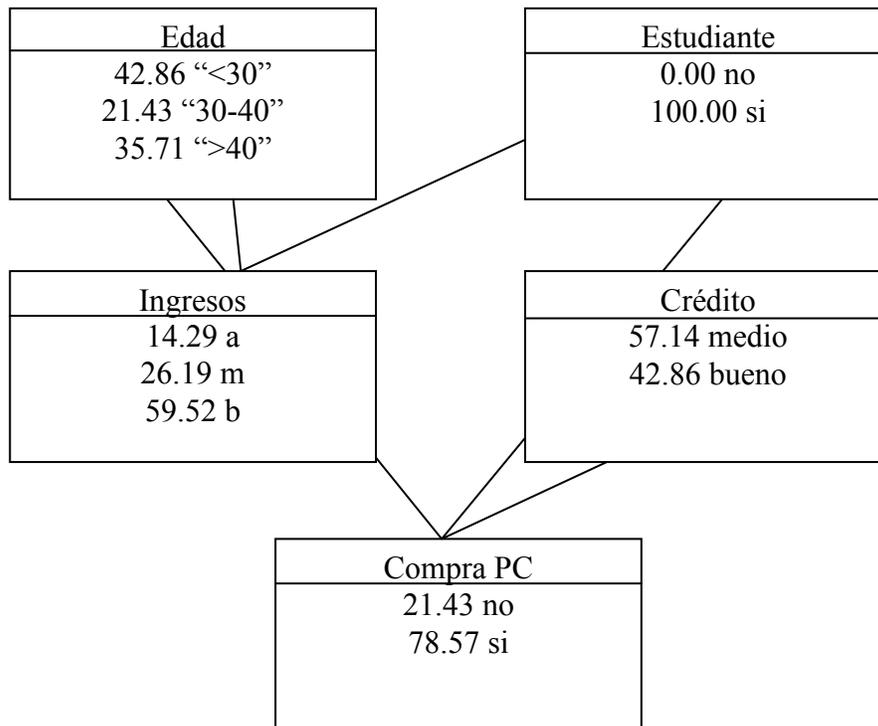


Figura 3.6 Probabilidad de compra de una PC, con evidencia de estudiante

3.1.1.5 Redes Neuronales

Las Redes Neuronales son sistemas conexionistas dentro del campo de la Inteligencia Artificial, las cuales, dependiendo del tipo de arquitectura neuronal, pueden tener diferentes aplicaciones. Pueden utilizarse en el reconocimiento de patrones, compresión de la información y reducción de dimensionalidad, agrupamiento, clasificación, etc.

Las Redes Neuronales tienen como finalidad, el emular los procesadores biológicos de información, por tanto para imitar esta característica debemos de basarnos en el uso de soportes artificiales semejantes a los que existen en el cerebro humano.

Las propiedades más interesantes del procesamiento neuronal humano son las siguientes:

- Procesamiento de información robusto y tolerante a fallos. Al inicio de la vida humana, el cerebro cuenta con el máximo número de neuronas. Desde ese momento, diariamente perdemos muchas miles de neuronas, y sin embargo, el

cerebro humano continúa funcionando durante muchos años sin alcanzar un deterioro que pueda afectar a nuestras capacidades.

- Procesadores de Información Flexibles. No necesitan volver a ser programados cuando se cambian de entorno o de ambiente, sino que ellos mismos se reajustan al entorno.
- Trabajan con información incompleta, ruidos o inconsistencias. Tienen la misma capacidad de procesamiento que las computadoras con programación muy sofisticada.

En la figura 3.7 se muestra el esquema de una Red Neuronal Artificial (RNA):

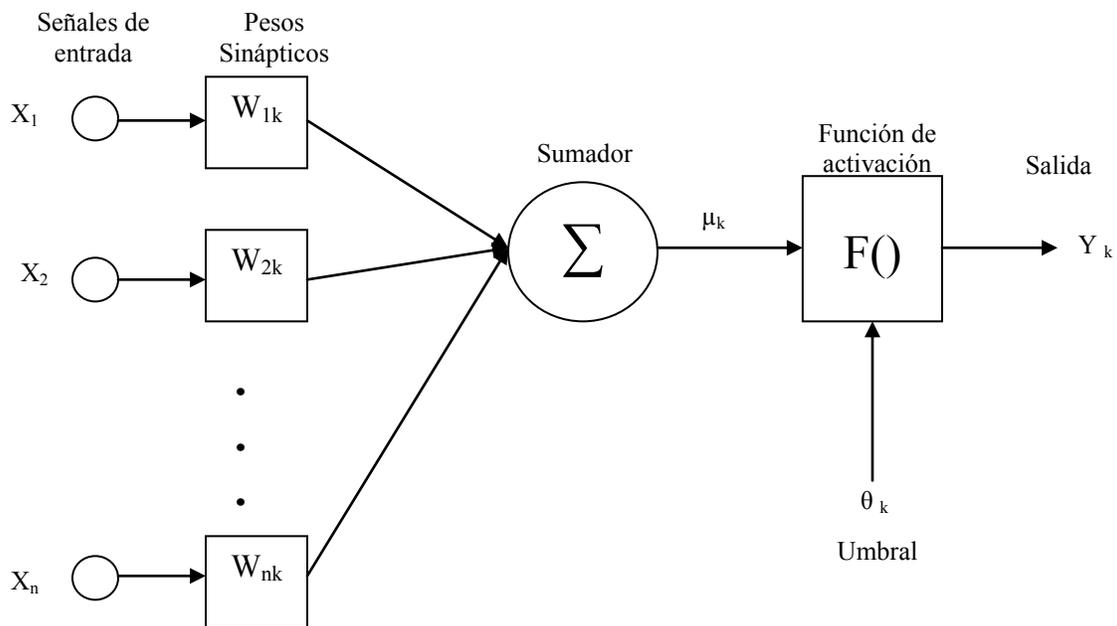


Figura 3.7 Esquema de una RNA

Las entradas a la neurona están representadas por el vector X , y el rendimiento de la sinapsis se modela mediante un vector de pesos W . En la figura 3.4 se muestra una entrada adicional. Entonces el valor de salida de esta neurona viene dado por:

$$y = f \left(\sum_i w_i x_i \right) = f(W \cdot X) = f(W^T X)$$

Donde f es la función de activación.

Cuando tenemos una red de neuronas, las salidas de unas se conectan con las entradas de otras. Si el peso de dos neuronas conectadas es positivo, el efecto producido es de excitación. Por el contrario, si es negativo, este efecto es de inhibición. Por tanto, una sola neurona es una unidad de procesamiento simple. Se considera que el potencial de las redes

neuronales artificiales proviene de la capacidad que proporciona el empleo de muchas de estas unidades simples y robustas al actuar en paralelo.

En la figura 3.9 se muestra un conjunto de entradas (el vector de entrada x) accediendo a la red desde el lado izquierdo y se propaga a través de la red hasta que la activación alcanza la capa de salida. Las capas intermedias son conocidas como capas ocultas ya que son invisibles desde fuera de la red.

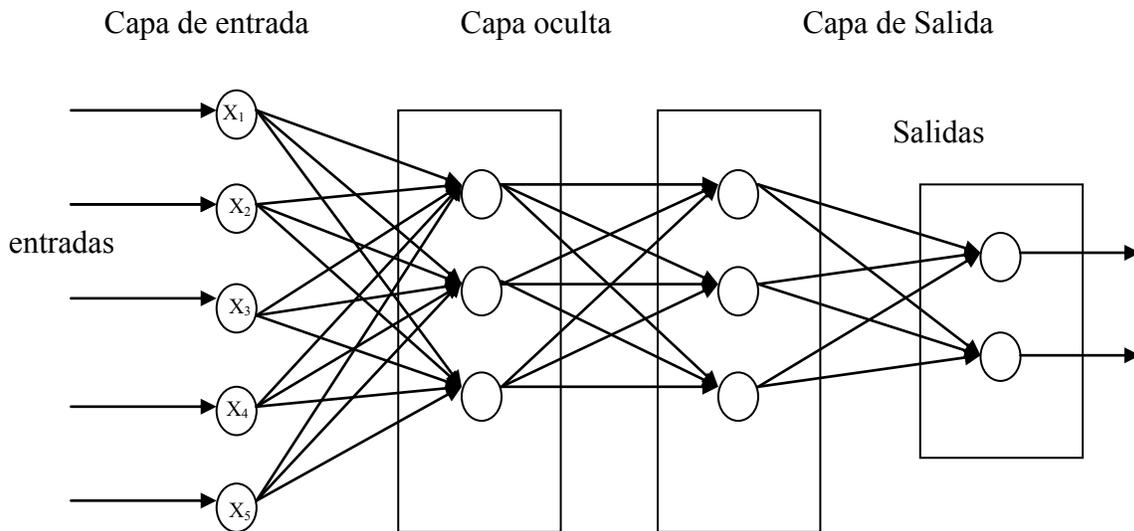


Figura 3.9 Red neuronal de tres capas

Hay dos modos de trabajo de una Red Neuronal Artificial:

- Modo de Transferencia de Activación: cuando la activación es transmitida por toda la red. Éste es el modo de funcionamiento o de aplicación y está asociado a la operación de propagación hacia adelante.
- Modo de aprendizaje: cuando la red se organiza normalmente a partir de la transferencia de activación más reciente.

3.1.2 Regresión

La regresión permite hacer predicciones de una variable X (*dependiente*) sobre otra variable Y (*independiente*) entre las que se intuye que existe una relación y calcula la relación de dependencia entre ellas; cuando solo existe una variable independiente se calcula la ecuación de la recta que mejor ajusta a los datos y se denomina regresión lineal y es la más utilizada. La ecuación de la recta tiene la siguiente forma:

$$y = mx + b$$

Donde m y b son variables que se calculan y que definen la posición e inclinación de la recta; donde m es la pendiente de la recta y b es la ordenada al origen. Para estimar los coeficientes a través de mínimos cuadrados se utilizan las siguientes fórmulas:

$$m = \frac{\sum xy - \bar{y} \sum x}{\sum x^2 - \bar{x} \sum x}$$

$$b = \bar{y} - m\bar{x}$$

Para determinar el coeficiente de determinación, es decir, que tanto se encuentran relacionadas las variables se utiliza la fórmula siguiente:

$$R = 1 - \frac{\sum e^2}{\sum (y - \bar{y})^2}$$

3.1.3 Reglas de Asociación

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.

3.1.3.1 Generación de Reglas

Las reglas de asociación surgieron inicialmente para afrontar el análisis de las cestas de la compra de los comercios; en este contexto, las diferentes cestas de la compra se pueden expresar formando una base de datos de una sola tabla, las filas de esta tabla se refieren a una cesta en un supermercado, mientras que las columnas son cada producto en venta en el supermercado. Un ejemplo se muestra en la tabla 3.2:

	Vino “El cabezón”	Gaseosa “Chispa”	Vino “Tío Paco”	Horchata “Xufer”	Bizcochos “Goloso”	Galletas “Trigo”	Chocolate “La vaca”
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1

Tabla 3.2 Tabla de la cesta de la compra

Un ejemplo de regla de asociación es el siguiente:

SI bizcochos “Golosos” **Y** horchata “Xufer” **ENTONCES** galletas “Trigo”

Por lo tanto una regla de asociación puede ser vista en su forma general como **SI α ENTONCES β** , donde α y β son dos conjuntos de elementos diferentes. Dada una regla de asociación, se suele trabajar con dos medidas para conocer la calidad de la regla: cobertura y confianza. La cobertura es el número de instancias que la regla predice correctamente, mientras que la confianza, mide el porcentaje de veces que la regla se cumple cuando se puede aplicar [3].

3.1.4 Agrupación

El agrupamiento o clustering es una de las tareas más frecuentes en la Minería de Datos; se trata de encontrar grupos entre un conjunto de individuos, por lo cual, el concepto de distancia puede jugar un papel crucial, ya que individuos similares (ceranos) deberían de ir para el mismo grupo. Muchos de los métodos de agrupamiento, se suelen incluir en el área que se utiliza en la estadística como análisis de conglomerados, aunque hoy en día este uso es más restringido debido a la hibridación⁷ de métodos.

3.1.4.1 Tipos de Agrupación

- Mapas auto-organizativos de Kohonen. Este método fue desarrollado por un finlandés llamado Teuvo Kohonen y en un principio fue desarrollado 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 [7].

3.1.4.2 Agrupación aglomerativa jerárquica

Uno de los principales problemas del agrupamiento, es discernir, cuántos grupos puede haber en los datos. Los métodos jerárquicos de agrupamiento se basan en la construcción de un árbol en el que las hojas son los elementos del conjunto de ejemplos, y el resto de los nodos son subconjuntos de ejemplos que pueden ser utilizados como particionamiento del espacio.

⁷ Es la combinación de dos o más métodos de minería de datos.

A este gráfico se le conoce como dendrograma y se muestra en la figura 3.9:

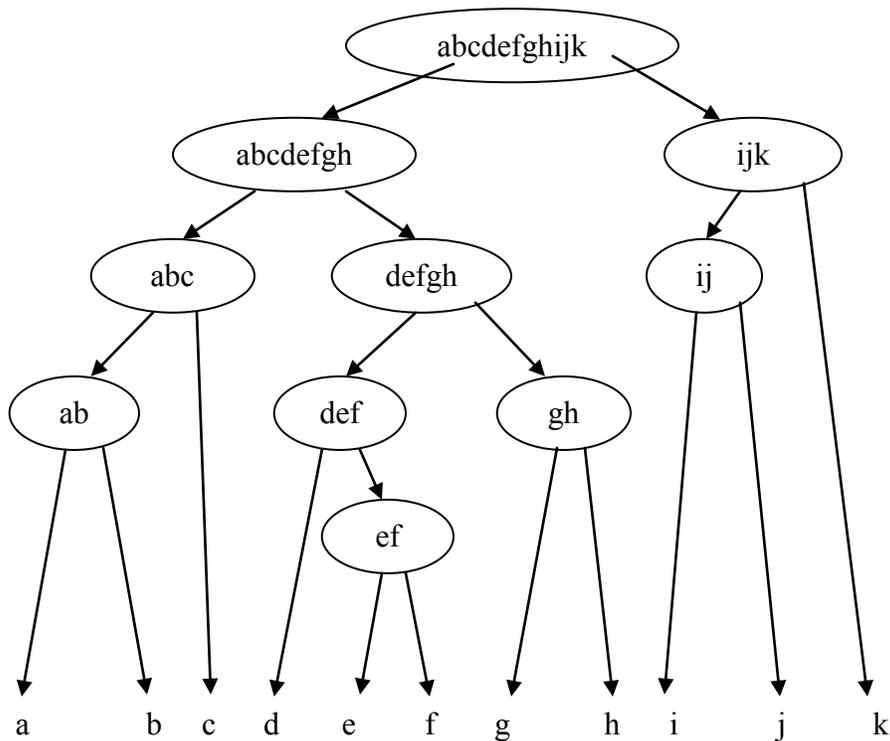


Figura 3.9 Ejemplo de un dendrograma

Una particularidad de este tipo de árbol, es que cada conjunto está situado en un nivel diferente de todos los demás. De esta forma se genera una jerarquía de conjuntos, que da nombre al conjunto de métodos, que permite la obtención de diferentes soluciones. Así en la figura 3.5 existe una jerarquía de 10 niveles:

1. Conjunto a, b, c, d, e, f, g, h, i, j, k
2. Conjunto a, b, c, d, e, f, g, h
3. Conjunto i, j, k
4. Conjunto i, j
5. Conjunto a, b, c
6. Conjunto d, e, f, g, h
7. Conjunto a, b
8. Conjunto g, h
9. Conjunto e, f

Esta estructura jerárquica permite generar varios agrupamientos, dependiendo de lo compacta que se desee la solución o del número de grupos a generar; para ello se elige un nivel en la jerarquía, se desprecian todos los descendientes de los nodos del mismo nivel y superior al seleccionado, y las hojas del árbol resultante definen el agrupamiento generado.

3.1.5 Correlación

La correlación indica la relación lineal que existe entre dos variables aleatorias. Se considera que dos variables aleatorias están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores de la otra. Por ejemplo, existe correlación entre dos variables A y B, si al aumentar los valores de A también aumentan los valores de B y viceversa [4].

Actualmente las empresas tienen la necesidad de calcular las correlaciones de las variables que afectan su mercado, por lo que esta tarea de minería de datos es muy popular y es utilizada también de forma empírica.

3.2 Algoritmos de Minería de Datos

Cada una de las tareas de Minería de Datos, requiere métodos y técnicas o algoritmos para resolverlas. Así como una tarea puede tener muchos métodos o algoritmos diferentes para resolverlas, tenemos que la misma tarea o al menos el mismo método puede resolver un gran abanico de técnicas o algoritmos. A continuación se muestra una tabla en la que podemos algunas tareas (clasificación, regresión, agrupamiento, reglas de asociación, etc.) y algunas técnicas o algoritmos que pueden abordarlas:

Nombre	Predictivo		Descriptivo	
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación
Redes Neuronales	✓	✓	✓	
Árboles de Decisión	✓	✓	✓	✓
Redes de Kohonen			✓	
Regresión Lineal y Logarítmica		✓		
Regresión Logística	✓			✓
Kmeans			✓	
A priori				✓
Naive Bayes	✓			
Vecinos más Próximos	✓	✓	✓	
Algoritmos genéticos y evolutivos	✓	✓	✓	

Tabla 3.3 Asociación de tareas y técnicas de Minería de Datos

3.2.1 Algoritmos Supervisados

Los algoritmos supervisados o predictivos, predicen el valor de una etiqueta o atributo de un conjunto de datos, 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; tal relación sirve para realizar la predicción en datos cuya etiqueta es desconocida.

Esta forma de trabajar se conoce como aprendizaje supervisado o predictivo. 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 continuos como son la regresión o las series temporales.

El aprendizaje supervisado requiere de una fase de entrenamiento en la cual se construye un modelo de predicción en el que se utiliza la mayor parte de los datos; también requiere de una fase de prueba en la que se prueba la validez del modelo probándolo con todos los datos.

3.2.2 Algoritmos no Supervisados

El aprendizaje no supervisado se caracteriza por descubrir modelos o características significativas a partir únicamente de los datos de entrada. Estos algoritmos Realizan tareas descriptivas como el descubrimiento de patrones y tendencias en los datos actuales. El descubrimiento de estos patrones sirve para llevar a cabo acciones y obtener un beneficio científico o de negocio de ellas.

CAPÍTULO IV

SOFTWARE LIBRE PARA LA IMPLEMENTACIÓN DE MINERÍA DE DATOS

4.1 WEKA (Waikato Environment for Knowledge Analysis)

Waikato Environment for Knowledge Analysis (WEKA), es un entorno para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos. Para ello únicamente se requiere que los datos a analizar se almacenen con un cierto formato, conocido como Attribute-Relation File Format (*ARFF*).

WEKA es un software de libre distribución desarrollado en Java. Está constituido por una serie de paquetes de código abierto con diferentes tareas como clasificación, agrupamiento, asociación, visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. Estos paquetes pueden ser integrados en cualquier proyecto de análisis de datos, e incluso pueden extenderse con contribuciones de los usuarios que desarrollen nuevos algoritmos. Con objeto de facilitar su uso por un mayor número de usuarios, WEKA además incluye una interfaz gráfica de usuario para acceder y configurar las diferentes herramientas integradas.

4.1.1 Historia

En 1993, la Universidad de Waikato en Nueva Zelanda, inició el desarrollo de la versión original de WEKA, y no fue hasta 1997 que se decidió reescribir su código en java incluyendo la implementación de nuevos algoritmos de modelado. En el 2005, WEKA recibe el galardón “Data Mining and Knowledge Discovery Services” (Servicios de Minería de Datos y Descubrimiento del Conocimiento).

4.1.2 Descripción

WEKA se distribuye como un archivo ejecutable comprimido de java (archivo "jar"), que se invoca directamente sobre la máquina virtual JVM. En las primeras versiones de WEKA se requería la máquina virtual Java 1.2 para invocar a la interfaz gráfica, desarrollada con el paquete gráfico de Java *Swing*.

En el caso de la última versión, WEKA 3-4, se requiere Java 1.3 o superior.

WEKA se distribuye como software de libre distribución desarrollado en Java. Está constituido por una serie de paquetes de código abierto con diferentes tareas como clasificación, agrupamiento, asociación, y visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. Estos paquetes pueden ser integrados en cualquier proyecto de análisis de datos, e incluso pueden extenderse con contribuciones de los usuarios que desarrollen nuevos algoritmos. Con objeto de facilitar su uso por un mayor número de usuarios, WEKA además incluye una interfaz gráfica de usuario para acceder y configurar las diferentes herramientas integradas.

4.1.2.1 Interfaz de Usuario

WEKA nos ofrece cuatro opciones posibles de trabajo:

- Simple CLI (Command Line Interface),
- Explorer
- Experimenter
- Knowledge Flow



Figura 4.1 Menú Principal de Weka

La opción de trabajo Simple CLI, es simplemente una ventana de comandos java para ejecutar las clases de WEKA. La primera distribución de WEKA no disponía de interfaz gráfica y las clases de sus paquetes se podían ejecutar desde la línea de comandos pasando los argumentos adecuados. En la figura 4.2 se muestra la pantalla de bienvenida de la interfaz de comandos de WEKA.

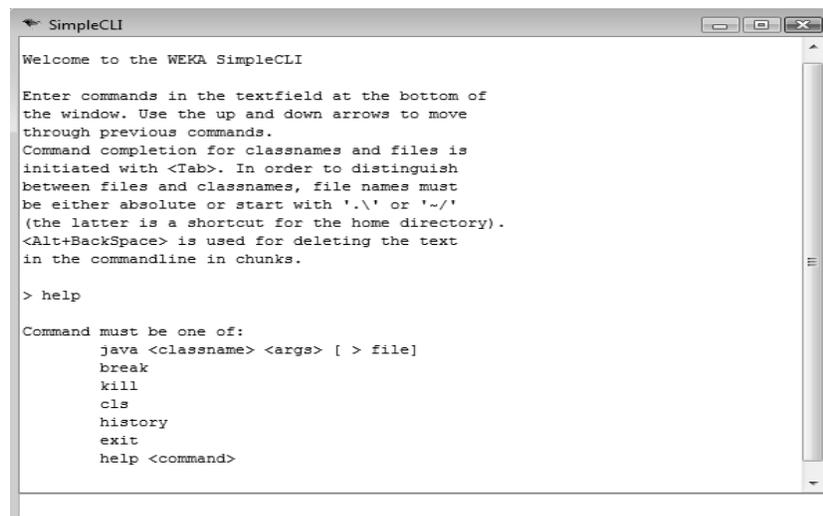


Figura 4.2 Ventana de Bienvenida Interfaz Simple CLI

4.1.2.1.1 Explorer

Esta opción de trabajo permite llevar a cabo la ejecución de los algoritmos de análisis implementados sobre los archivos de entrada, una ejecución independiente por cada prueba. En la figura 4.3 se observa la pantalla de la ventana principal de la opción de trabajo Explorer de WEKA.

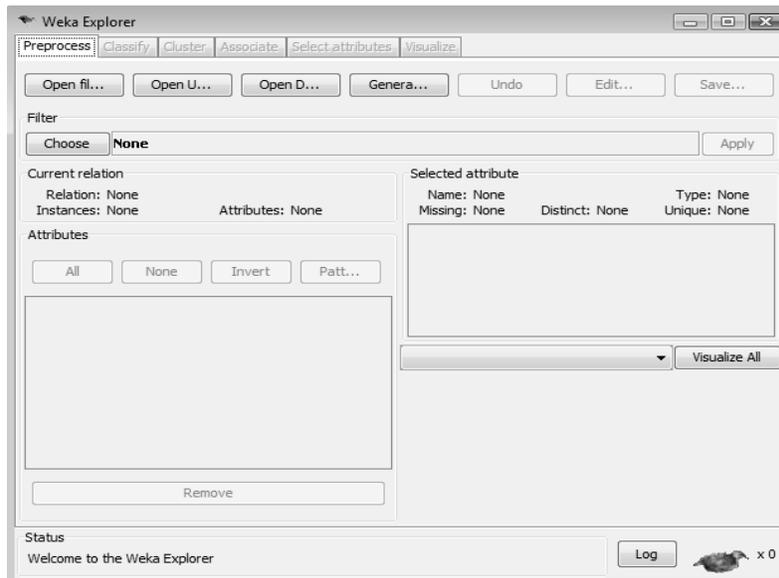


Figura 4.3 Ventana Principal Opción Explorer de Weka

El primer paso para comenzar a trabajar con el explorador es definir el origen de los datos. Weka soporta diferentes fuentes que coinciden con los botones que están debajo de las pestañas superiores mostrados en la figura 4.3. Las diferentes posibilidades son las siguientes:

1. Open File

Al pulsar sobre este botón aparecerá una ventana de selección de archivo. Aunque el formato por defecto de Weka es el arff eso no significa que sea el único que admita, para ello tiene intérpretes de otros formatos. Otros formatos que Weka admite son los siguientes:

- CSV Archivos separados por comas o tabuladores. La primera línea contiene los atributos.
- C4.5 Archivos codificados según el formato C4.5. Unos datos codificados según este formato estarían agrupados de tal manera que en una búsqueda, en un archivo .names estarían los nombres de los atributos y en un archivo .data estarían los datos en sí. Weka cuando lee archivos codificados según el formato C4.5 asume que ambos archivos (el de definición de atributos y el de datos) están en el mismo directorio, por lo que sólo es necesario especificar uno de los dos.

- Instancias Serializadas Weka internamente almacena cada muestra de los datos como una instancia de la clase instance. Esta clase es serializable por lo que estos objetos pueden ser volcados directamente sobre un archivo y también cargados de uno.

Para cargar un archivo arff simplemente debemos buscar la ruta donde se encuentra el archivo y seleccionarlo. Si dicho archivo no tiene extensión arff, al abrirlo Weka intentará interpretarlo, si no lo consigue aparecerá un mensaje de error.

2. Open Url

Con este botón se abrirá una ventana que nos permitirá introducir una dirección en la que definir dónde se encuentra nuestro archivo. El tratamiento de los archivos (restricciones de formato, etc.) es el mismo que el apartado anterior.

3. Open DB

Con este botón se nos da la posibilidad de obtener los datos de una base de datos. Para configurarlo, lo primero que hay que hacer es definir la url desde la cual se conectará a la base de datos, la contraseña para acceder, el nombre de usuario y la consulta que se desea realizar y si deseamos o no usar el modo de datos abreviado. En la figura 4.4 se muestra la pantalla de conexión de OpenDB.

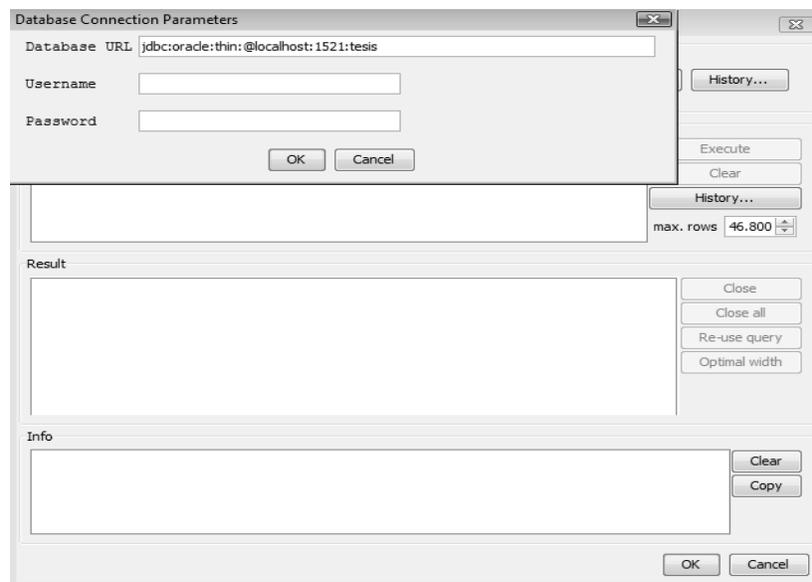
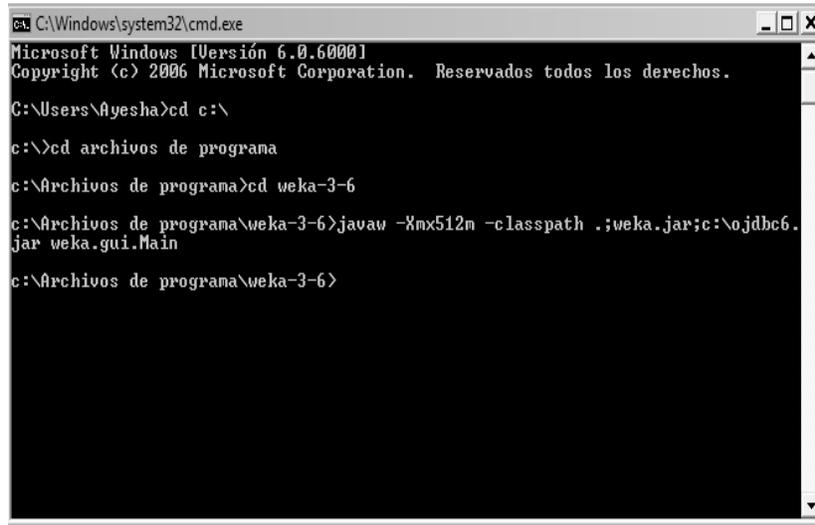


Figura 4.4 Ventana de conexión OpenDB

Al utilizar este modo de obtención de datos podemos realizar las consultas desde la interfaz de WEKA realizando de esta forma minería de datos relacional o consultar una vista minable sin necesidad de crear el archivo arff.

Para poder conectar WEKA con el manejador de bases de datos, en este caso oracle, es necesario contar con el jdbc correspondiente a la versión del manejador que se está

utilizando e inicializar WEKA mediante un comando específico de java como se muestra en la figura 4.5.



```
C:\Windows\system32\cmd.exe
Microsoft Windows [Versión 6.0.6000]
Copyright (c) 2006 Microsoft Corporation. Reservados todos los derechos.

C:\Users\Ayesha>cd c:\
c:\>cd archivos de programa
c:\Archivos de programa>cd weka-3-6
c:\Archivos de programa\weka-3-6>javaw -Xmx512m -classpath .;weka.jar;c:\ojdbc6.jar weka.gui.Main
c:\Archivos de programa\weka-3-6>
```

Figura 4.5 Inicializando WEKA con OpenDB

WEKA se inicializa mostrando la pantalla de bienvenida de la figura 4.6.

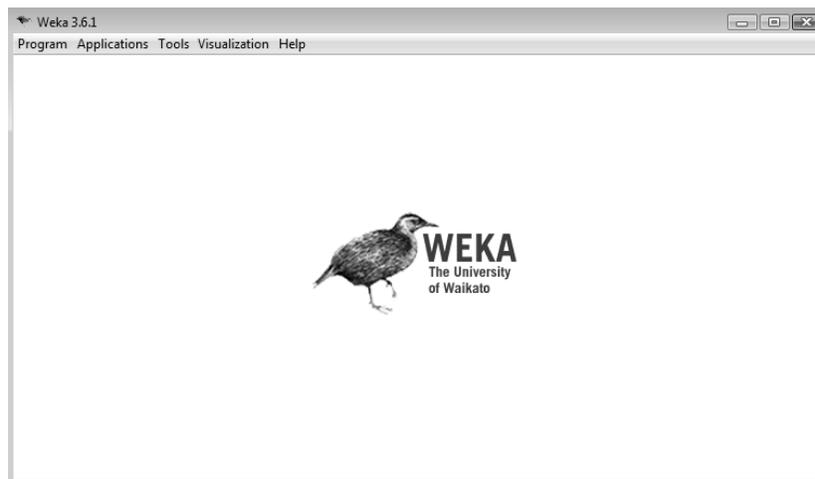


Figura 4.6 Pantalla de Bienvenida de WEKA con OpenDB

Una vez realizada esta secuencia de acciones podemos seleccionar de la opción Applications la forma de Explorer de WEKA y seleccionar OpenDB para trabajar directamente con la base de datos para ello seleccionamos “Applications” y después “Explorer”: En la figura 4.7 se puede observar la ventana principal de la opción Explorer abierta desde consola de Windows.

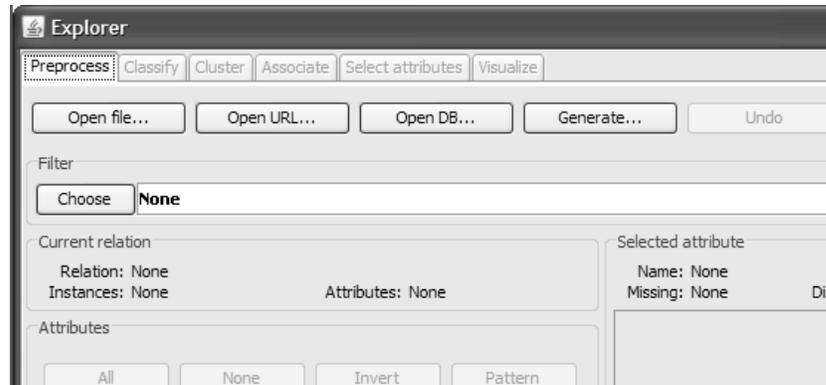


Figura 4.7 Explorer con OpenDB

Podemos seleccionar la opción “Open DB...” notaremos que el campo “URL” se encuentra con la dirección de la base de datos utilizada. En el botón “User...” especificamos el nombre de usuario y contraseña con el que se conectará a la base de datos y posteriormente se da click en el botón Connect. Finalmente Weka consigue el acceso y en el campo “Query” se puede escribir la consulta que se desea realizar a la base de datos. En la figura 4.8 se observa el resultado de la ejecución de una consulta realizada desde OpenDB.

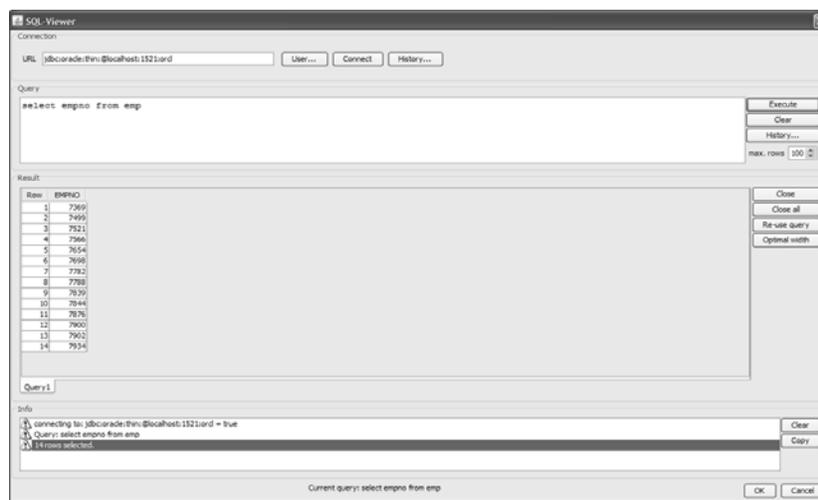


Figura 4.8 Weka en OpenDB

Después de haber elaborado los pasos anteriores (obtener datos directamente de la base de datos) lo único que se tiene que hacer es dar click en “OK” de la misma ventana y Weka regresará a la pantalla inicial de Explorer pero con los datos de las columnas seleccionadas de la base de datos.

Una vez ubicados en esta ventana podremos aplicar cualquier método adecuado para los datos seleccionados como se muestra en las figuras 4.9 y 4.10.

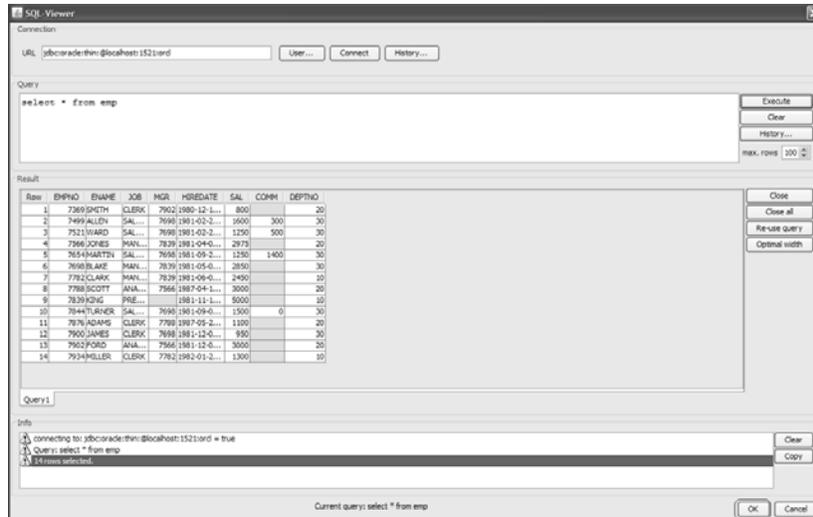


Figura 4.9 Resultado de Consulta OpenDB

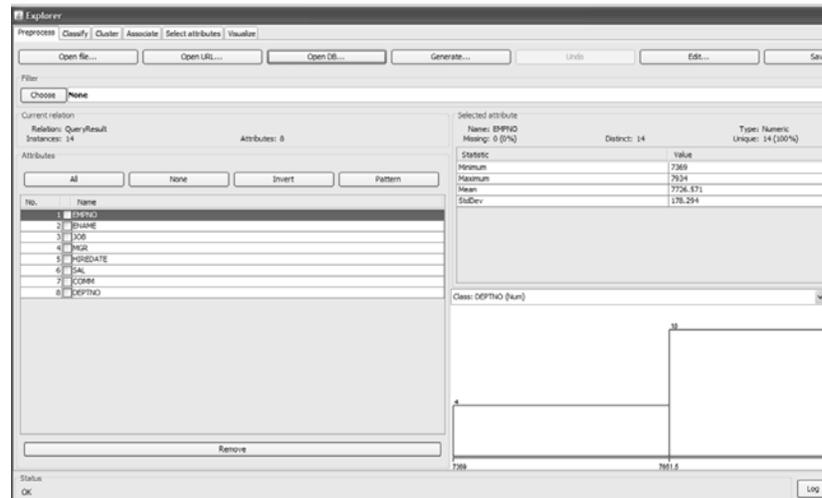


Figura 4.10 Visualización de Datos obtenidos desde Oracle

4.1.2.1.2 Experimenter

Esta opción permite definir experimentos más complejos, con objeto de ejecutar uno o varios algoritmos sobre uno o varios conjuntos de datos de entrada, y comparar estadísticamente los resultados. El modo experimentador (*Experimenter*) es un modo muy útil para aplicar uno o varios métodos de clasificación sobre un gran conjunto de datos y, luego poder realizar contrastes estadísticos entre ellos y obtener otros índices estadísticos. En la figura 4.11 se puede observar la ventana principal de la opción experimenter.

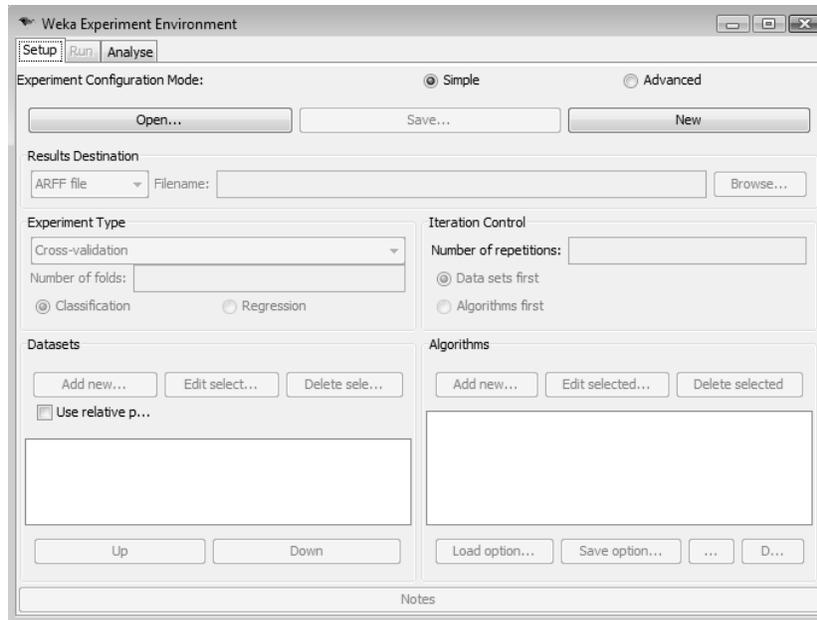


Figura 4.11 Ventana Principal Opción Experimenter de Weka

4.1.2.1.3 Knowledge Flow

Esta opción es una novedad de WEKA que permite llevar a cabo las mismas acciones de la Interfaz Explorer, con una configuración totalmente gráfica, inspirada en herramientas de tipo "data-flow" para seleccionar componentes y conectarlos en un proyecto de minería de datos, desde que se cargan los datos, se aplican algoritmos de tratamiento y análisis, hasta el tipo de evaluación deseada.

Esta última opción de Weka es quizá la más cuidada y la que muestra de una forma más explícita el funcionamiento interno del programa. Su funcionamiento es gráfico y se basa en situar en el panel de trabajo, elementos base de manera que creamos un "circuito" que defina nuestro experimento. A continuación se muestra en la figura 4.12 la ventana principal de la opción Knowledge Flow.

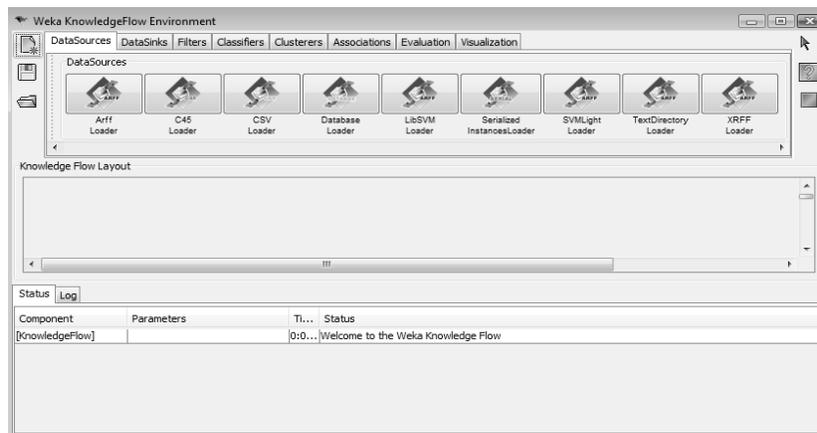


Figura 4.12 Ventana Principal Interfaz Knowledge Flow de Weka

4.1.3 Características de WEKA

WEKA es una extensa colección de algoritmos de Máquinas de conocimiento desarrollados por la universidad de Waikato (Nueva Zelanda) implementados en Java [2]; útiles para ser aplicados sobre datos mediante las interfaces que ofrece o para embeberlos dentro de cualquier aplicación. Además WEKA contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. WEKA está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla.

Sin embargo, y pese a todas las cualidades que WEKA posee, tiene un gran defecto y éste es la escasa documentación orientada al usuario que conlleva a una utilización bastante pobre, lo que la hace una herramienta difícil de comprender y manejar sin información adicional.

La licencia de WEKA es GPL*, lo que significa que este programa es de libre distribución y difusión. Además, ya que WEKA está programado en Java, es independiente de la arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible.

4.2 Rapid Miner

RapidMiner es un software de distribución libre usado más comúnmente por las empresas, a diferencia de WEKA que es un software para aprendizaje sobre Minería de Datos, por lo cual es considerado por algunos especialistas como el líder mundial de código abierto en minería de datos; debido a la combinación de tecnología de primera calidad y su rango de funcionalidad. Además de ser una herramienta flexible para aprender y explorar la minería de datos, la interfaz gráfica de usuario tiene como objetivo simplificar el uso para las tareas complejas de esta área.

4.2.1 Historia

La versión inicial fue desarrollada por el departamento de inteligencia artificial de la Universidad de Dortmund en 2001. Se distribuye bajo licencia GPL y está hospedado en SourceForge desde el 2004.

RapidMiner proporciona más de 500 operadores orientados al análisis de datos, incluyendo los necesarios para realizar operaciones de entrada y salida, preprocesamiento de datos y visualización.

4.2.2 Descripción

RapidMiner (anteriormente, YALE, Yet Another Learning Environment) es un programa informático para máquinas de aprendizaje y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. El concepto modular de Rapidminer permite que el operador diseñe y anide un gran número de cadenas para complejos problemas de aprendizaje, debido a que el

manejo interno de datos es transparente para el usuario; RapidMiner introduce nuevos conceptos de transparencia del manejo de datos y modelado de procesos, en donde la configuración final del proceso está dada por el usuario.

4.2.2.1 Usuario

RapidMiner se puede utilizar para tareas de máquinas de aprendizaje, procesos de descubrimiento de conocimiento, y para un gran número de temas que incluye la generación y selección de características, conceptos de manejo de datos, preprocesamientos de datos, agrupamiento, clasificación y procesamiento de textos.

4.2.3 Características de Rapid Miner

Su aplicación más común es en la investigación, y algunas de sus características principales son las siguientes:

- Se encuentra desarrollado en Java
- Es un sistema multiplataforma
- Representación interna de los procesos de análisis de datos en archivos XML
- Permite el desarrollo de programas a través de un lenguaje de script
- Puede usarse de diversas maneras:
 - A través de un GUI
 - En línea de comandos
 - En batch
 - Desde otros programas a través de llamadas a sus librerías
- Es extensible
- Incluye gráficos y herramientas de visualización de datos.

CAPÍTULO V

REVISIÓN DE ENCUESTA DE PROFESORES AL FINAL DEL SEMESTRE EN LA FACULTAD DE INGENIERÍA REALIZADA POR EL ING. JOSÉ ENRIQUE LARIOS CANALE.

5.1 Introducción

En el siguiente capítulo se hablará sobre un análisis de la encuesta de profesores, que se realiza cada fin de semestre a los alumnos de la Facultad de Ingeniería, mediante la cual, se obtienen datos estadísticos sobre la calidad del proceso enseñanza-aprendizaje, en la Facultad de Ingeniería.

Para llevar a cabo el estudio de minería de datos se seleccionaron algunas preguntas de la encuesta, sin embargo no todas coinciden con la validez de estudio realizado en este capítulo; las preguntas seleccionadas que coinciden en su validez son únicamente 2 y 4.

En los últimos años, muchos países han sido testigos de reformas y transformaciones en sus sistemas de educación superior, incluyendo la aparición de nuevos tipos de instituciones, cambios en los patrones de financiamiento y gobierno, establecimiento de mecanismos de evaluación, acreditación y reformas curriculares e innovaciones tecnológicas. Sin embargo, el panorama es heterogéneo y complejo, la educación superior en México no está cambiando con la misma rapidez y existen múltiples agendas, tanto institucionales como locales y regionales, por lo que cada institución y sistema de educación superior tendrá que tomar decisiones que les permitan adaptarse a las nuevas realidades que día a día exigen su transformación.

La educación superior en México enfrenta varios desafíos, uno de ellos es transformarse a efecto de ser parte de la sociedad mundial del conocimiento y la información; dicha transformación tendrá que contar con un eje basado en una visión innovadora y un nuevo paradigma para la formación de profesionistas. Tal paradigma deberá incluir, entre otros elementos, la educación a lo largo de la vida (learning long life), el aprendizaje autodirigido, la formación integral con una visión humanista y la responsabilidad ante el desarrollo sustentable.

Los desafíos que enfrenta la Facultad de Ingeniería (FI) de la Universidad Nacional Autónoma de México (UNAM) son 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 en la FI 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 cuenta la FI para cumplir con

sus objetivos educativos. Dicho esfuerzo debe sustentarse en el convencimiento personal, en el compromiso de cada uno de los miembros de la institución, especialmente de los alumnos y de los profesores.

En este contexto, mejorar las estrategias de enseñanza que desarrolla el profesor redundarán en una mejora al proceso enseñanza-aprendizaje. 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.

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.

5.2 El contexto mundial y la educación superior

Existen tres grandes cambios vinculados a la transformación de la educación superior:

- 1) La globalización económica vista como un proceso creciente de integración de capitales, tecnología e información a través de las fronteras nacionales, lo que genera un mercado mundial integrado que consecuentemente obliga a los países a competir en una economía global.
- 2) La importancia creciente del conocimiento en el desarrollo económico se encuentra cada vez más ligado a la habilidad de una nación de adquirir o aplicar conocimiento con base en las innovaciones tecnológicas y del uso competitivo del conocimiento.
- 3) La revolución de las comunicaciones y la información ha transformado radicalmente la capacidad de almacenar, transmitir y utilizar la información. Las innovaciones en electrónica y telecomunicaciones, así como el desarrollo de tecnología para aumentar la capacidad de transmisión de información, aunado a la reducción de sus costos, ha tenido como consecuencia la eliminación de las distancias físicas; gracias a un mayor acceso a fuentes de información y al establecimiento de comunicaciones entre personas, instituciones y países que cuentan con servicios de internet.

Las economías sustentadas en el conocimiento requieren de sus trabajadores un mayor nivel de habilidades. Esto ha sido ilustrado en información reciente sobre tasas de retorno en algunos países de América Latina (Chile, Brasil y México), las cuales muestran una creciente tasa de retorno por la educación terciaria.

En virtud de la segunda dimensión del cambio a la que hemos hecho alusión, referida al progresivo rol del conocimiento, adquiere creciente importancia la educación continua, dada la necesidad de actualización constante en conocimientos y habilidades. El enfoque tradicional de estudiar una vez por todas para obtener un título o para concluir un posgrado, antes de iniciar o continuar con el desarrollo profesional, está siendo reemplazado por prácticas de educación a lo largo de la vida. A fin de satisfacer estas demandas educativas son necesarios una serie de programas a impulsarse en la FI, que van desde la consolidación de los cuerpos académicos, el impulso y desarrollo de la innovación educativa, la vinculación interinstitucional, la construcción de un nuevo perfil en la práctica de la gestión, la planeación y la evaluación institucional, en la que destaca de forma especial la evaluación del proceso enseñanza-aprendizaje y, particularmente, la evaluación del profesor.

5.3 Planteamiento del Problema

Con el fin de mejorar la calidad de la educación en México, se han creado instancias a nivel nacional para evaluar a las instituciones educativas de todos los niveles. La certificación expedida por las instancias evaluadoras califican la calidad o excelencia de los programas educativos que se imparten en las instituciones de educación, tanto públicas como privadas. La Facultad de Ingeniería, como el resto de las dependencias docentes de la UNAM, ha sido objeto de estas evaluaciones tanto en sus programas a nivel licenciatura como de posgrado.

Las evaluaciones hechas a la FI han abarcado algunos temas sobre la profesionalización de la labor docente del personal académico con parámetros que no han sido considerados en los diferentes programas de superación que se le ofrecen al profesorado de la FI, sobre todo no se le ha preparado 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 talento debe ser una de las tareas que la FI debe llevar a cabo en su proceso de selección y promoción del profesorado.

En la FI se ha considerado que la revisión de los planes y programas de estudio, los ajustes y cambios en la dirección y administración de la educación, las mejoras en la planta física, mejorarán la calidad de la educación, sin reforzar adecuadamente el planteamiento de la metodología educativa que se necesita aplicar al plan de estudios diseñado para las carreras que se imparten en la FI y desarrollar los correspondientes programas para la formación de profesores y su profesionalización.

El plantear las estrategias educativas para la profesionalización del profesorado en la FI conlleva la necesidad de evaluarlo en su quehacer docente, por lo que inicialmente se analizarán los mecanismos actuales de evaluación que se aplican al personal académico de

la FI, tanto los establecidos en el marco jurídico previsto por la legislación universitaria, como los mecanismos y los correspondientes instrumentos que la administración de la FI ha desarrollado con esta finalidad.

El análisis de los mecanismos de evaluación que actualmente se aplican al profesorado de la FI proporcionará un marco de referencia sobre el cual se sustente la propuesta de evaluación del profesorado que se presenta en este documento, que se fundamenta sobre una investigación educativa de los esfuerzos que se han desarrollado en instituciones educativas nacionales y de otros países sobre este tema, ajustándola a las necesidades y características propias de los planes y programas de estudio de la FI de la UNAM.

5.4 Marco Jurídico Universitario

De los 47 apartados de la Legislación Universitaria, dos apartados se refieren exclusivamente al personal académico y a la labor docente: el ESTATUTO DEL PERSONAL ACADÉMICO DE LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO (EPA) y el MARCO INSTITUCIONAL DE DOCENCIA.

En el Artículo 1º del EPA se establece que este ESTATUTO rige las relaciones entre la Universidad y su personal académico de acuerdo a la LEY ORGÁNICA y el ESTATUTO GENERAL DE LA UNAM. Haciendo un breve análisis del articulado de cada capítulo del EPA, de los diferentes temas que plantea sobre el profesorado, sólo en dos capítulos se trata el tema de la evaluación: en el CAPÍTULO II De los Concursos de Oposición o Concursos Abiertos establece los mecanismos de evaluación del profesional que aspira a ingresar a la planta laboral de la UNAM en calidad de personal académico para lo cual en este CAPÍTULO se indican los parámetros a evaluar para su selección; sobre este tema, también el CAPÍTULO III De los Concursos de Oposición Para Promoción o Concursos Cerrados establece que, después de tres años de actividad académica, el profesor puede solicitar que se abra este tipo de concurso para promoverse a otra categoría o nivel lo cual conllevan una evaluación de los méritos adquiridos por el docente en este periodo de tiempo acorde a lo estipulado por el EPA.

En ambos casos, concursos abiertos y cerrados, el EPA establece los lineamientos generales sobre los cuales se evalúa a los participantes. Las dependencias académicas son las que emiten la convocatoria en la que se establecen los términos en que se evaluará a los que soliciten participar en el caso de los concursos abiertos, y el académico que desee promoverse solicitará al Consejo Técnico respectivo se abra el concurso cerrado. Para los dos tipos de concursos el Consejo Técnico, en el caso de la FI, ha integrado comisiones dictaminadoras por División de adscripción que apoyadas en comisiones auxiliares llevan a cabo el proceso de evaluación.

Analizando los criterios de evaluación o valoración como se indica en el EPA, se puede observar que los aspectos pedagógicos, sobre comunicación, interacción social,

psicopedagógicos, no son explicitados o su evaluación es limitada. De hecho, además de evaluar el *currículum vita*, del profesor, principalmente se evalúan conocimientos sobre el área de la ingeniería, para los profesores de carrera, o la asignatura, si es para profesor de asignatura; mediante un interrogatorio que llevan a cabo los integrantes de la comisión auxiliar.

También se evalúa el conocimiento que el aspirante tenga sobre los programas de las asignaturas correspondientes al área de conocimientos, si se concursa para profesor de carrera, o de la asignatura en cuestión si es para profesor de la misma, mediante una prueba escrita en la que el aspirante desarrolla una crítica sobre el o los programas que la comisión auxiliar le solicite.

Por otra parte, la comisión auxiliar califica la exposición de un tema que le asigna al participante, esta exposición es ante un grupo de alumnos, ya sean de su grupo, si ya da clases, o de otro maestro. Si el aspirante ya desempeña actividades docentes, la comisión dictaminadora las evalúa en términos de asistencia, puntualidad, colaboración, resultados de las encuestas semestrales, etc. En el EPA no están considerados procedimientos o mecanismos continuos y permanentes para evaluar el desempeño de las actividades docentes del personal académico de la Institución. Tampoco están consideradas comisiones u órganos internos, salvo los mencionados para los concursos abiertos y cerrados.

El Marco Institucional de Docencia es un documento que contiene un conjunto de principios y conceptos en los que se basa el quehacer docente de la Universidad, basados sobre la misión de la Institución, que le da esencia a estas actividades. Este conjunto de enunciados establecen los principios en que se basa la docencia pero no se refieren directamente a actividades concretas de evaluación del docente.

5.5 La Evaluación del Docente en la FI

5.5.1 Antecedentes

Hace 30 años, aproximadamente, un grupo de profesores de matemáticas, desarrolló un instrumento de evaluación del desempeño académico del profesorado de la División de Ciencias Básicas (DCB) de la FI. Este instrumento lo diseñaron para ser contestado por los alumnos del profesor a quien se deseara evaluar y se planteó que previo a la terminación de clases, es decir, entre la tercera y la segunda semana antes de concluir el ciclo escolar, las autoridades académico-administrativas instrumentaran su aplicación para que a la hora de clase se le proporcionara a los alumnos el cuestionario que contestarían en un lapso de tiempo de 15 a 20 minutos. Se les indicó a los alumnos que el instrumento de evaluación se le proporcionaría al profesor hasta el siguiente semestre, esto con el fin de que los alumnos lo contestaran abiertamente, sin reticencias, ya que, se les dijo, que su profesor se enteraría de cómo lo evaluaron hasta el siguiente semestre.

En ese entonces, las autoridades de la FI, previo a la aplicación de la encuesta a la comunidad de profesores, puntualizaron que los resultados de la encuesta serían de

conocimiento exclusivo del profesor y que de ninguna manera afectaría su historial académico, es decir, que los resultados de la encuesta no se asentarían en su expediente, ni se utilizarían como parámetros de evaluación en los concursos de oposición o promoción.

Con estos argumentos, en una asamblea de la Unión de Profesores de la FI, la representación del profesorado manifestó la aceptación para la aplicación de la encuesta, en la confianza de que el profesor del grupo efectivamente recibiría de su autoridad los instrumentos de evaluación contestados por los alumnos tal como se habían guardado en el sobre después de haber sido contestados por los alumnos, respetando las firmas con las que se selló, para que él abriera el sobre y evaluara las respuestas de sus alumnos en dichos instrumentos, al inicio del siguiente semestre.

Posteriormente, a fines del semestre en que se tuvo la mencionada asamblea de la Unión de Profesores, este instrumento de evaluación del docente se aplicó a todos los profesores de la FI y se les explicó que dicho instrumento de evaluación era una encuesta que sus alumnos contestarían para observarle los aspectos en los que el profesor podía mejorar su labor docente, con lo cual incidiría directamente en el proceso enseñanza-aprendizaje y por consiguiente mejorar el sistema educativo de la Facultad. Desde hace unos 25 años a la fecha, cuando se acerca el fin de semestre, coloquialmente las autoridades académico-administrativas comentan con los profesores que ya se va aplicar la encuesta.

5.5.2 Análisis del Instrumento

A continuación se hace un análisis del instrumento que contestan los alumnos sobre su profesor, análisis que se desglosa en diferentes aspectos con el fin de responder a preguntas tales como ¿Cuál es la naturaleza y finalidad del instrumento? ¿En qué ámbito del sistema educativo se encuadran las preguntas? ¿Cuál es la efectividad del instrumento?, entre otras.

5.5.2.1 ¿Evaluación o encuesta?

Para establecer la naturaleza del instrumento se revisara lo que el mismo instrumento establece, para lo cual se reproduce el encabezado y la parte final del mismo, que a la letra dice:

ENCABEZADO

FACULTAD DE INGENIERÍA
PROGRAMA DE EVALUACIÓN DEL
PROCESO ENSEÑANZA-APRENDIZAJE

Opinión sobre el desempeño académico del profesor y del alumno

Nombre del profesor: _____

Asignatura: _____ Semestre: _____

Estimado alumno:

Con objeto de mejorar la calidad educativa de nuestra Facultad periódicamente se realiza la evaluación del proceso enseñanza-aprendizaje. Dentro de este programa el punto de vista de los alumnos, acerca del ejercicio docente y de su propio desempeño como estudiante, constituye un factor muy importante.

Analice cada uno de los enunciados y llene la respuesta que considere más adecuada (una sola opción por pregunta). Le agradecemos contestar todas las preguntas.

PARTE FINAL DEL INSTRUMENTO

Le agradezco su disposición para contestar esta encuesta.

Atentamente

EL DIRECTOR.

Cabe mencionar que días previos a la aplicación del instrumento, se publica un cartel dirigido a los profesores solicitándoles su colaboración para llevar a cabo la evaluación de su actividad docente. Este cartel lo firma el Secretario de Servicios Académicos de la FI.

De la sola lectura del cartel que publica la Secretaria de Servicios Académicos, del encabezado y de la parte final del instrumento no se puede precisar qué es y para qué sirve, para ¿evaluar el proceso enseñanza-aprendizaje? ¿Evaluar al docente? ¿Tener la opinión o el punto de vista del alumno sobre el docente y sobre el mismo? o ¿Es una encuesta sobre el profesor?. La primera observación que se hace a este instrumento está precisamente en lo ambiguo que por sí mismo se define. Y en consecuencia lo que se pretende con el mismo.

Aunque es obvia la diferencia entre los términos evaluación y encuesta, no está por demás señalarlo: según la Real Academia de la Lengua Española. *Encuesta es un acopio de datos obtenidos mediante consulta o interrogatorio, referentes a estados de opinión, costumbres, nivel económico o cualquier otro aspecto de la actividad humana.* De la misma fuente, se establece que evaluación es sinónimo de valoración o valuación que define al *acto de valorar o señalar a una cosa el valor que le corresponde a su estimación.* Lo anterior evidencia que la evaluación especifica la definición de un valor, en nuestro caso a la labor docente, y en el caso de la encuesta sería el acopio de datos de opinión sobre el profesor.

Si se quiere llegar a una conclusión sobre cuál es el objetivo de aplicar este instrumento al profesorado de la Facultad, en principio se puede pensar más en que es un instrumento de evaluación, al señalarse en el mismo que es una acción del Programa de Evaluación. Además, porque el procesamiento del instrumento lo realiza la Facultad y los resultados se entregan a las autoridades académico-administrativas, con una copia al expediente del profesor para que los resultados se tomen en cuenta tanto en participaciones del profesor en concursos abiertos y cerrados, como para aumentar o disminuir sus responsabilidades académicas.

Al profesor se le regresa el sobre abierto con las “encuestas” y dos hojas con el resumen de los resultados: una gráfica y un análisis comparativo de sus datos con respecto del resto de profesores de la asignatura.

5.5.2.2 Análisis de las Preguntas del Instrumento

El instrumento tiene un total de 29 preguntas organizadas en dos bloques: SOBRE EL PROFESOR correspondiéndole 20 preguntas y las restantes a SOBRE EL ALUMNO (Autoevaluación). A continuación se analiza el primer bloque de preguntas, de acuerdo a lo planteado en el punto anterior, considerando que las preguntas 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 de este bloque aporta información que efectivamente evalúe al docente como tal, tomando en consideración que la respuesta no es abierta sino de opción múltiple (cuatro opciones).

a) SOBRE EL PROFESOR

- 1. El interés del profesor para que los alumnos aprendan es: Muy grande; Grande; Escaso, y Nulo.*

No es procedente la pregunta, ya que 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, quien puede evaluar si el profesor tiene o no interés en el aprendizaje del alumno.

- 2. La confianza que el profesor inspira a los alumnos para intervenir en clase es: Muy grande; Grande; Escaso, y Nulo.*

Si es procedente la pregunta, ya que el profesor puede o no inspirar confianza al alumno para intervenir en clase.

- 3. Las clases se desarrollan de manera: Muy interesante; Interesante; Aburrida, y Muy aburrida.*

No es procedente la pregunta, ya que no está bien dirigida la pregunta, pues debe encauzarse 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. Esta evaluación es competencia de un especialista en el tema.

- 4. El profesor expone la clase en forma clara: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si es procedente la pregunta, ya que el alumno puede percibir la capacidad didáctica de su profesor para lograr una exposición clara.

5. *Los conocimientos del profesor sobre la asignatura son : Excelentes; Buenos; Regulares; Deficientes.*

No es procedente la pregunta, el alumno no tiene las competencias para valorar los conocimientos que posee su profesor sobre la asignatura. Esta evaluación es competencia de una autoridad académica.

6. *El profesor propicia que los alumnos amplíen por su cuenta sus conocimientos sobre la asignatura: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si es procedente la pregunta, ya que el alumno puede percatarse si el profesor propicia que él desarrolle sus propias habilidades.

7. *El profesor propicia que los alumnos razonen sobre los temas que expone: Siempre; Frecuentemente; Algunas veces; Nunca.*

No es procedente la pregunta, el alumno no tiene las competencias para valorar los conocimientos psicopedagógicos que posee su profesor. Esta evaluación es competencia de un especialista en el tema.

8. *El profesor promueve un ambiente de respeto en clase: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si procede la pregunta, ya que el alumno es capaz de valorar el ambiente grupal que genera el profesor.

9. *Las tareas, trabajos y/o ejercicios que deja el profesor contribuyen al aprendizaje: Siempre; Frecuentemente; Algunas veces; Nunca.*

No es procedente la pregunta, el alumno no puede saber si las tareas, trabajos y/o ejercicios que deja el profesor contribuyen a su aprendizaje, pues no tiene los conocimientos sobre la asignatura. Esta evaluación sobre el profesor es competencia de una autoridad académica.

10. *Al exponer los temas el profesor fomenta la habilidad de plantear y resolver problemas: Siempre; Frecuentemente; Algunas veces; Nunca.*

No es procedente la pregunta, el alumno no tiene las competencias para valorar los conocimientos psicopedagógicos que posee su profesor. Esta evaluación es competencia de un especialista en el tema.

11. *El profesor llega generalmente al salón de clase: A la hora señalada; Con un retraso de 10 a 20 minutos; Con un retraso de 21 a 30 minutos; Con un retraso de mas de 30 minutos.*

No es procedente la pregunta, el control de la puntualidad 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.

12. *El profesor regresa los exámenes, tareas, trabajos y proyectos corregidos: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo. La evaluación de este tema es también competencia de una autoridad académica.

13. *Los exámenes y/o evaluaciones tienen un grado de dificultad adecuado: Siempre; Frecuentemente; Algunas veces; Nunca.*

No es procedente la pregunta, el alumno no tiene los conocimientos sobre la asignatura para valorar el grado de dificultad de los exámenes y/o evaluaciones sean colegiados o del profesor, por ejemplo, si un alumno no estudia cualquier problema le parecerá que tiene un alto grado de dificultad. Esta evaluación es competencia de una autoridad académica.

14. *El profesor termina su clase: A la hora señalada; Antes de la hora señalada; Entre 10 y 5 minutos antes; Más de 10 minutos antes.*

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.

15. *El profesor entrega oportunamente las calificaciones: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si es procedente la pregunta, ya que el alumno puede retroalimentar al profesor sobre el cumplimiento de sus responsabilidades ante el grupo. La evaluación de este tema es también competencia de una autoridad académica.

16. *El profesor faltó a clase durante el semestre: Nunca; De 1 a 3 veces; De 4 a 6 veces; Más de 6 veces.*

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.

17. *La clase se da sin perder de vista los puntos a donde se quiere llegar: Siempre; Frecuentemente; Algunas veces; Nunca.*

No es procedente la pregunta, se debe redactar en términos de los objetivos para esa clase, además, el alumno no tiene los conocimientos sobre la asignatura para valorar si se cumple con los objetivos planteados por el profesor. Esta evaluación es competencia de una autoridad académica.

18. *Al definir principios y conceptos se presentan ejemplos que faciliten la comprensión: Siempre; Frecuentemente; Algunas veces; Nunca.*

Si es procedente la pregunta, ya que el alumno puede percibir los apoyos didácticos a los que recurre el profesor para mejorar su aprendizaje mediante la resolución de ejemplos o ejercicios que resuelva el alumno con la orientación de su profesor.

19. *En caso de tener la oportunidad ¿volvería a tomar clase con el profesor? Definitivamente si; Probablemente si; Probablemente no; Definitivamente no.*

Si es procedente la pregunta, ya que el alumno muestra tener elementos para valorar a su profesor positivamente, un elemento fundamental puede ser la el aprendizaje obtenido por el alumno en la signatura que cursa con su profesor.

20. *Al inicio del curso se dieron a conocer los objetivos: Si; No.*

Si es procedente la pregunta, ya que al alumno le permite conocer los alcances del curso en cuanto a carga de estudio y ser motivador para sus expectativas de conocimientos por adquirir.

Del análisis anteriormente expuesto se concluye que 8 de las 20 preguntas hechas a los alumnos permiten evaluar competencias del docente. Sin embargo, hay aspectos relevantes del desempeño académico del profesor cuyos indicadores no es posible que sean evaluados por el alumno, como es el caso de las otras 12 preguntas. Por otra parte, el instrumento de evaluación no establece en qué orden ni cuáles son las competencias que se están evaluando y faltan temas que deben ser considerados en este instrumento.

b) SOBRE EL ALUMNO (Autoevaluación)

21. Falto a clase durante el semestre: Nunca; De 1 a 3 veces; De 4 a 6 veces; más de 6 veces.
22. Por lo general al asistir a clase llego: A la hora señalad; Con un retraso de 10 a 20 minutos; Con un retraso de 21 a 30 minutos; Con un retraso de más de 30 minutos.
23. Para esta asignatura, estudio por mi cuenta a la semana: Más de 1 hora; Entre 1 y 2 horas; Entre 2 y 3 horas; Mas de 3 horas.
24. Participo activamente en clase (preguntando, poniendo atención aportando ejemplos, etc.): Siempre; Frecuentemente; Algunas veces; Nunca.
25. Me siento motivado para estudiar la asignatura: Siempre; Frecuentemente; Algunas veces; Nunca.
26. Mi aprendizaje de los temas de la asignatura ha sido: Excelente; Bueno; Regular; Deficiente.
27. Realizo las tareas y trabajos que me asigna el profesor: Siempre; Frecuentemente; Algunas veces; Nunca.
28. La actitud del grupo ayudó a mejorar mi desempeño: Siempre; Frecuentemente; Algunas veces; Nunca.
29. Mi condición académica en esta asignatura es: Primera Inscripción; Segunda Inscripción; Sin derecho a inscripción ARTÍCULO 27; Sin derecho a inscripción ARTÍCULO 19.

De las nueve preguntas de la autoevaluación del alumno, de la 21 a la 28 son más bien preguntas de reflexión que, bajo la guía u orientación del profesor, puede hacer el grupo dos o tres veces al semestre, por ejemplo, después de conocer los resultados de una evaluación, con la finalidad de analizar cómo se está llevando a cabo el proceso enseñanza-aprendizaje. Con respecto a la pregunta 29, aparentemente no tiene sentido, pues la información la conoce la autoridad administrativa y el alumno, y para el profesor no le sirve, ya que la conoce cuando el estudiante dejó de ser su alumno.

En cuanto a los “Comentarios y sugerencias al profesor respecto a la impartición de la asignatura.”, al ser una pregunta abierta, en ocasiones los comentarios son de mucha ayuda para que el profesor conozca sobre su labor docente. Un porcentaje del 30 % no suele hacer comentario alguno. Al ser anónima la evaluación, en ocasiones los comentarios son irrelevantes o inadecuados.

CAPÍTULO VI ANTECEDENTES DE LA MINERÍA DE DATOS A REALIZARSE

6.1 Estructura de la Encuesta de Profesores

La encuesta de profesores al final del semestre de la Facultad de Ingeniería se encuentra conformada por una serie de preguntas, a su vez divididas en dos partes, sobre el profesor y sobre el alumno. Las preguntas del profesor se encuentran divididas en cuatro conceptos importantes del proceso de enseñanza-aprendizaje: aprendizaje, motivación, evaluación y responsabilidad; mientras que las preguntas relacionadas con el alumno se dividen sólo en dos conceptos: responsabilidad del alumno y desempeño del alumno. Cada uno de los conceptos, tanto del profesor como del alumno, son evaluados y su calificación aparece en la hoja de resultados de las encuestas.

En la figura 6.1 se muestra la estructura básica de la encuesta, en donde se puede observar que las preguntas se dividen en dos grupos principales: alumno y profesor; también se puede observar que la mayoría de las preguntas están agrupadas por las características del profesor o del alumno sobre las cuales se hace referencia, a estas características a lo largo de esta tesis se les llamarán conceptos. En cuanto a las preguntas del profesor, la encuesta propone 4 conceptos:

1. Aprendizaje
2. Motivación
3. Evaluación
4. Responsabilidad

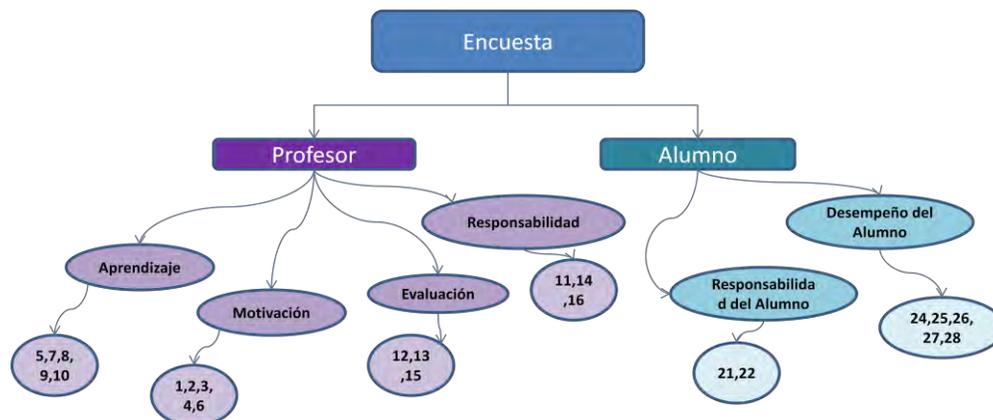


Figura 6.1 Estructura general de la encuesta

En la tabla 6.1 se muestra una correlación entre algunas preguntas y conceptos, del profesor y del alumno, en las que se observa una posible relación de efectos de un resultado en una pregunta sobre el profesor y el desempeño del alumno; en las siguientes correlaciones se basará el proceso de minería de datos del capítulo 7.

Profesor	Alumno	Conceptos	
Confianza que inspira a los alumnos para intervenir en clase (P02)	Participo activamente en clase(P24)	Motivación	Desempeño del Alumno
Propicia que los alumnos amplíen por su cuenta los conocimientos sobre la asignatura(P07)	Me siento motivado para estudiar la asignatura(P25)	Motivación	Desempeño del Alumno
Los conocimientos del profesor sobre la asignatura son(P05)	Mi aprendizaje de los temas de la asignatura ha sido(P26)	Aprendizaje	Desempeño del Alumno
Asistencia a clase(P16)	Realizo tareas y trabajos asignados(P27)	Responsabilidad	Desempeño del Alumno

Tabla 6.1 Correlación de preguntas seleccionadas entre alumno y profesor

En la tabla anterior se puede observar que la relación entre la calificación obtenida en una pregunta del profesor y en una pregunta del alumno, puede estar reflejada en una calificación de concepto tanto del alumno como del profesor. Por lo tanto, el análisis se puede llevar a cabo tanto desde el punto de vista de las preguntas como desde el punto de vista de los conceptos de enseñanza-aprendizaje que son manejados en la encuesta.

En la figura 6.2 se observa la ubicación en los conceptos de las preguntas contenidas en la tabla 6.1.

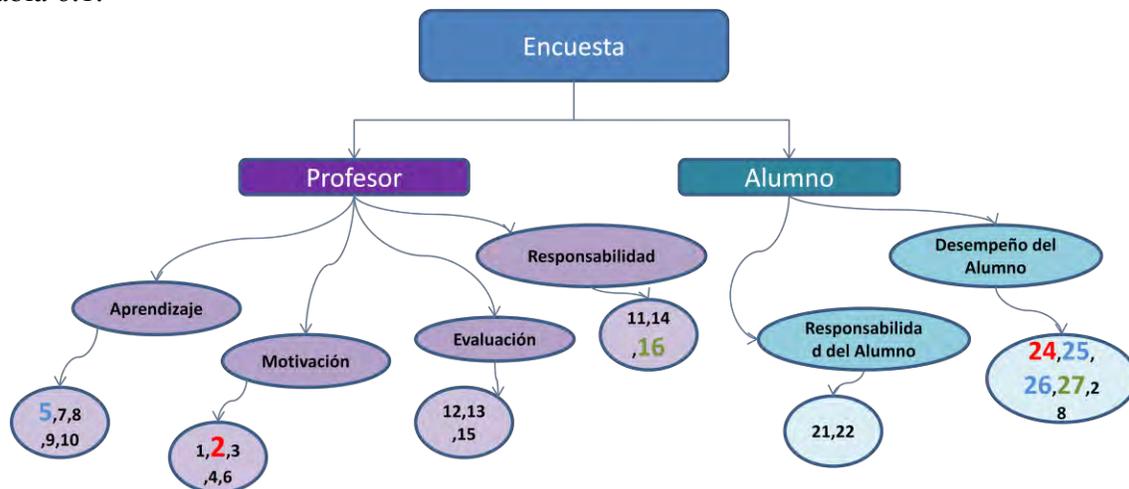


Figura 6.2 Esquema de correlación de preguntas y conceptos tabla 6.3

En la tabla 6.2 se muestra un análisis semejante al anterior pero éste sólo maneja preguntas y conceptos del alumno, con el fin de encontrar un punto de vista más objetivo sobre los resultados de las encuestas de los profesores; en la figura 6.3 se muestran la ubicación en los conceptos de las preguntas contenidas en la tabla 6.2.

	Alumno		Conceptos
Falté a clases durante el semestre(P21)	Mi aprendizaje de los temas ha sido(P26)	Asistencia	Desempeño del Alumno
Me siento motivado para estudiar la asignatura(P25)	La actitud del grupo ayudó a mejorar mi desempeño(P28)	Motivación	Desempeño del Alumno
Participo en clase(P24)	Realizo tareas y trabajos asignados(P27)	Participación	Desempeño del Alumno

Tabla 6.2 Correlación entre preguntas seleccionadas sobre el alumno

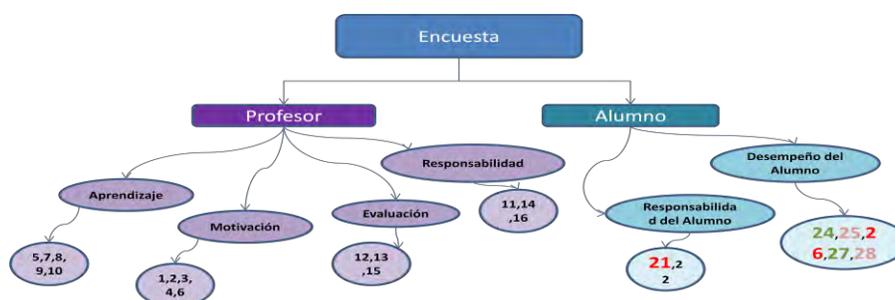


Figura 6.3 Esquema de correlación entre preguntas y conceptos tabla 6.2

En la tabla 6.3 se muestra una correlación entre preguntas y conceptos únicamente sobre el profesor, en las primeras dos columnas se muestran las preguntas sobre el profesor que se encuentran relacionadas; en las segundas dos columnas se muestran los conceptos que se encuentran involucrados en dichas preguntas; en la figura 6.4 se muestra la ubicación en los conceptos de las preguntas utilizadas en la tabla 6.3.

Profesor			Conceptos
El profesor expone en forma clara(P04)	Los conocimientos del profesor sobre la asignatura son(P05)	Motivación	Aprendizaje
El interés del profesor por que los alumnos aprendan es(P01)	Volvería a tomar clases con el profesor(P19)	Motivación	
El profesor propicia que los alumnos razonen sobre los temas que expone(P07)	El profesor fomenta la habilidad de plantear y resolver problemas(P10)	Aprendizaje	Aprendizaje
Las tareas, trabajos y/o ejercicios que deja el profesor contribuyen al aprendizaje(P09)	Adecuado grado de dificultad en evaluaciones y/o exámenes(P13)	Aprendizaje	Evaluación

Tabla 6.3 Correlación entre preguntas seleccionadas sobre el profesor

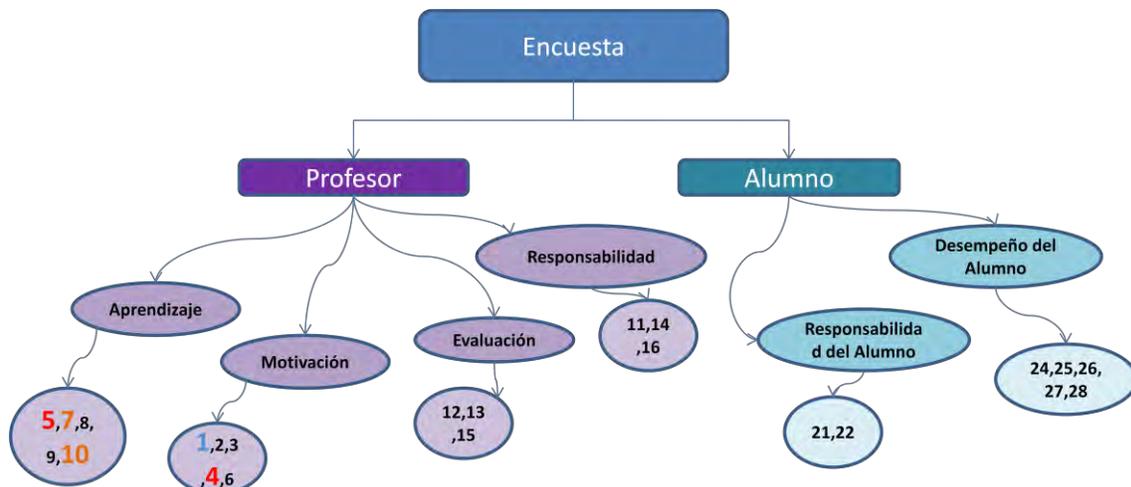


Figura 6.4 Esquema de correlación entre preguntas y conceptos tabla 6.3

Al analizar la relación entre las preguntas del alumno y el profesor, sólo del alumno y sólo del profesor podemos obtener una visión más objetiva sobre los resultados que obtuvo un profesor en un grupo y de esta forma lograr una evaluación completa sobre el proceso de enseñanza-aprendizaje. Se puede observar que los conceptos engloban algunas habilidades del profesor y que además, obtiene una calificación sobre ellas. Debido a eso, se puede obtener una correlación entre las calificaciones de concepto y la calificación general del profesor, y de esta forma indicar cuáles son las habilidades que predominan en su clase y cuáles no, y además que influyen directamente en su calificación general.

En la figura 6.5 se muestra la correlación entre las calificaciones de concepto y la calificación general del profesor.



Figura 6.5 Relación entre conceptos y calificación general

CAPÍTULO VII

DESARROLLO DE MINERÍA DE DATOS PARA LA BASE DE DATOS DE LAS ENCUESTAS DE PROFESORES DE LA FACULTAD DE INGENIERÍA

7.1 Documentación de la Base de Datos de Encuestas de Profesores

La base de datos de encuestas de profesores utilizada para desarrollo de esta tesis fue proporcionada por la Secretaría de Servicios Académicos de la Facultad de Ingeniería [12]. Dicha base de datos no cuenta con los elementos reales respecto a los RFC y nombres de Profesores, debido a que, con el fin de preservar la confidencialidad de los resultados de las encuestas los nombres de profesores y RFC no corresponden con sus respectivos resultados de encuestas; de tal forma que el análisis de las encuestas se lleva a cabo respecto a los nombres de las asignaturas, cuyos elementos y resultados no fueron modificados. La base de datos tiene información desde el semestre 1998-1 hasta el semestre 2010-1.

La encuesta de profesores al final del semestre, se encuentra compuesta de las siguientes preguntas:

Profesor

1. El interés del profesor para que los alumnos aprendan es
2. La confianza que el profesor inspira a los alumnos para intervenir en clase es
3. Las clases se desarrollan de manera interesante
4. El profesor expone con claridad la clase
5. Los conocimientos del profesor sobre la asignatura son
6. El profesor propicia que los alumnos amplíen por su cuenta sus conocimientos
7. El profesor propicia que los alumnos razonen sobre los temas que expone
8. El profesor promueve un ambiente de respeto en clase
9. Las tareas, trabajos y/o ejercicios que deja el profesor, contribuyen al aprendizaje
10. Al exponer los temas, el profesor fomenta la habilidad de plantear y resolver problemas
11. El profesor llega generalmente al salón de clase
12. El profesor regresa los exámenes, tareas, trabajos y proyectos corregidos
13. Los exámenes y/o evaluaciones tienen un grado de dificultad Adecuado
14. El profesor termina su clase
15. El profesor entrega oportunamente las calificaciones
16. El profesor faltó a clase durante el semestre
17. La clase sin perder de vista los puntos a donde se quiere llegar
18. Al definir principios o conceptos se presentan ejemplos que facilitan la comprensión
19. En caso de tener la oportunidad ¿volvería a tomar clase con el profesor?
20. Al inicio del curso se dieron a conocer los objetivos

Alumno

21. Falté a clases durante el semestre
22. Por lo general al asistir a clases llego
23. Para esta asignatura estudio por mi cuenta a la semana
24. Participo activamente en clase(preguntando, poniendo atención, aportando ejemplos, etc.)

25. Me siento motivado para estudiar la asignatura
26. Mi aprendizaje de los temas de la asignatura ha sido
27. Realizo tareas y trabajos que me asigna el profesor
28. La actitud del grupo ayudó a mejorar mi desempeño
29. Mi condición académica en esta asignatura es

Evaluación de Conceptos

1. Aprendizaje (5, 7, 8, 9, 10)
2. Motivación (1, 2, 3, 4, 6)
3. Evaluación (12, 13, 15)
4. Responsabilidad (11, 14, 16)
5. Desempeño del alumno (24, 25, 26, 27, 28)
6. Responsabilidad del alumno (21, 22)

Debido a las nuevas herramientas que nos proporciona software especializado como WEKA y RapidMiner, surge la necesidad de contar con la información de la base de datos, no solamente en Excel, sino también en un gestor de bases de datos; para esta tesis se eligió Oracle versión 11g como gestor para la base datos.

La identificación de entidades y atributos para la creación de la base de datos, se llevó a cabo en base a la información proporcionada que se compone de las preguntas y conceptos desarrollados durante el apartado 7.1 del presente capítulo; además de la información sobre la encuesta, también fue proporcionada información adicional sobre ella, a continuación se lista los atributos adicionales proporcionados:

- Clave de División
- Clave de Departamento
- Clave de Asignatura
- Nombre de la Asignatura
- Grupo
- Semestre
- Tipo de asignatura(teoría o laboratorio)

Cabe señalar que la información proporcionada para esta tesis fue entregada en un archivo de Excel, pero fue exportada de una base de datos de la Secretaría de Servicios Académicos por lo que después de la identificación de las entidades atributos se procedió a realizar la carga de los datos a través de SQL Loader y con la creación de archivos ejecutables que se encargan de llamar a los archivos de control y de datos correspondientes a cada tabla. El archivo de control tiene la siguiente estructura:

```
load data
infile "asignatura.csv" badfile "asignatura.bat" discardfile "asignatura.dsc"
insert
into table ASIGNATURA
FIELDS TERMINATED BY ','TRAILING NULLCOLS
(NUM_ASIGNATURA,
NOM_ASIGNATURA,
TIPO_ASIGNATURA);
```

Donde los nombres separados por comas que se encuentran dentro de los paréntesis hacen referencia a los atributos de la tabla y estos deben de estar en el orden en que se encuentran en el archivo de Excel. El archivo de Excel debe de ser de tipo *.CSV que se refiere a un archivo con datos separados por comas.

7.1.1 Diagrama Físico Relacional

En la figura 7.1 se muestra el diagrama físico del modelo relacional, el cual consiste en mostrar gráficamente las entidades, sus atributos y características principales y la relación que existe entre las entidades. Este diagrama corresponde a la base de datos proporcionada por la Secretaría de Servicios Académicos de la Facultad de Ingeniería [12].

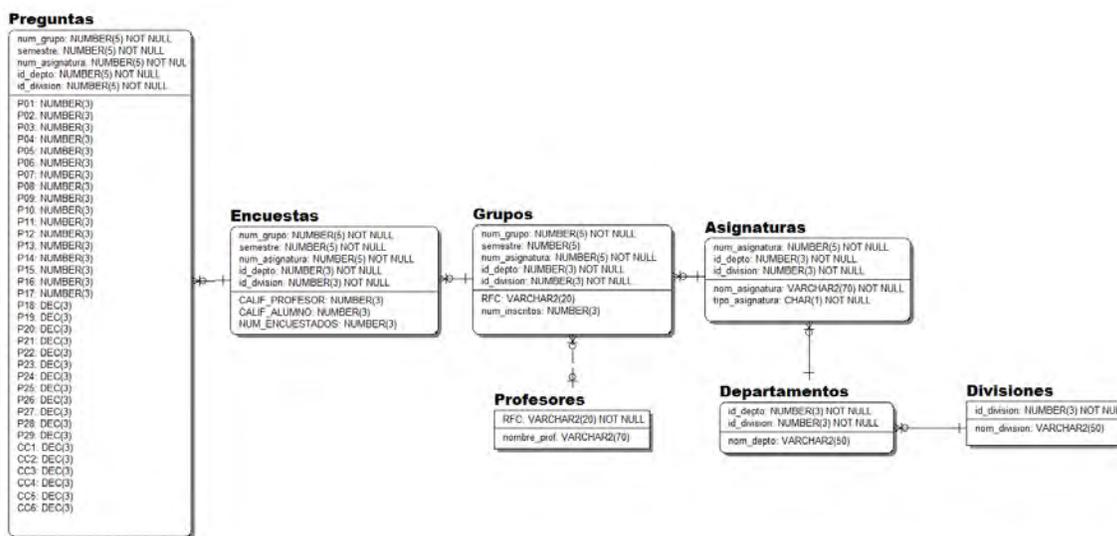


Figura 7.1 Diagrama Entidad Relación de Encuestas de Profesores

7.1.2 Diccionario de Datos Corporativo

El diccionario de datos Corporativo es un conjunto de metadatos⁸ que contiene las características lógicas y puntuales de los datos que se encuentran dentro de la Base de Datos; debe contener la información de los detalles y descripción del problema. En este apartado se describen las tablas que conforman la base de datos proporcionada por la Secretaría de Servicios Académicos de la Facultad de Ingeniería.

⁸ Información sobre los datos, e.j. tamaño, tipo descripción, etc.

En la tabla 7.1 se muestra la información contenida en la tabla llamada División, la cual almacena la información proporcionada sobre las divisiones.

Divisiones

Nombre	Atributo	Descripción	Obligatoriedad
ID_DIVISION	Identificador de división	NUMBER(4)	PK NN
NOM_DIVISION	Nombre de División	VARCHAR2(50)	N

Tabla 7.1 Datos de División

En la tabla 7.2 se muestra la información contenida en la tabla llamada Departamento, la cual almacena la información proporcionada sobre los departamentos y la llave primaria de la tabla de división.

Departamentos

Nombre	Atributo	Descripción	Obligatoriedad
ID_DEPTO	Identificador de departamento	NUMBER(4)	PK NN
ID_DIVISION	Identificador de división	NUMBER(4)	FK ₁ NN
NOM_DEPTO	Nombre de departamento	VARCHAR2(50)	N

Tabla 7.2 Datos de Departamento

En la tabla 7.3 se muestra la información contenida en la tabla llamada Asignatura, la cual almacena la información proporcionada sobre las asignaturas y la llave primaria de la tabla de su correspondiente división y departamento.

Asignaturas

Nombre	Atributo	Descripción	Obligatoriedad
NUM_ASIGNATURA	Número de Asignatura	NUMBER(4)	PK NN
ID_DEPTO	Identificador de departamento	NUMBER(4)	FK ₂ NN
ID_DIVISION	Identificador de división	NUMBER(4)	FK ₁ NN
NOM_ASIGNATURA	Nombre de Asignatura	VARCHAR2(60)	N
TIPO_ASIGNATURA	Tipo de Asignatura	CHAR(1)	N

Tabla 7.3 Datos de Asignatura

En la tabla 7.4 se muestra la información contenida en la tabla llamada Grupos, la cual almacena la información proporcionada sobre los grupos y la llave primaria de la tabla de su correspondiente división y departamento.

Grupos

Nombre	Atributo	Descripción	Obligatoriedad
NUM_GRUPO	Número de Grupo	NUMBER(4)	PK NN
NUM_ASIGNATURA	Número de Asignatura	NUMBER(4)	FK ₃ NN
ID_DEPTO	Identificador de departamento	NUMBER(4)	FK ₂ NN
ID_DIVISION	Identificador de división	NUMBER(4)	FK ₁ NN
NUM_INSCRITOS	Número de inscritos	NUMBER(2)	N

Tabla 7.4 Datos de Asignatura

En la tabla 7.5 se muestra la información contenida en la tabla llamada Encuestas, la cual almacena la información proporcionada sobre los datos principales de las encuestas.

Encuestas

Nombre	Atributo	Descripción	Obligatoriedad
NOM_ASIGNATURA	Nombre de Asignatura	VARCHAR2(60)	PK NN
SEMESTRE	Semestre	NUMBER(4)	PK NN
NUM_GRUPO	Número de Grupo	NUMBER(4)	FK ₅ NN
ID_DEPTO	Identificador de departamento	NUMBER(4)	FK ₂ NN
ID_DIVISION	Identificador de división	NUMBER(4)	FK ₁ NN
CALIF_ALUMNO	Calificación del Alumno	NUMBER(4)	N
CALIF_PROFESOR	Calificación del Profesor	NUMBER(3)	N
NUM_ENCUESTADOS	Número de encuestados	NUMBER(3)	N

Tabla 7.5 Datos de Encuesta

En la tabla 7.6 se muestra la información contenida en la tabla llamada Preguntas, la cual almacena las calificaciones sobre las preguntas de la encuesta y conceptos.

Preguntas

Nombre	Atributo	Descripción	Obligatoriedad
P01	Pregunta 01	NUMBER(3)	N
P02	Pregunta 02	NUMBER(3)	N
P03	Pregunta 03	NUMBER(3)	N
P04	Pregunta 04	NUMBER(3)	N
P05	Pregunta 05	NUMBER(3)	N
P06	Pregunta 06	NUMBER(3)	N
P07	Pregunta 07	NUMBER(3)	N
P08	Pregunta 08	NUMBER(3)	N
P09	Pregunta 09	NUMBER(3)	N
P10	Pregunta 10	NUMBER(3)	N
P11	Pregunta 11	NUMBER(3)	N
P12	Pregunta 12	NUMBER(3)	N
P13	Pregunta 13	NUMBER(3)	N
P14	Pregunta 14	NUMBER(3)	N
P15	Pregunta 15	NUMBER(3)	N
P16	Pregunta 16	NUMBER(3)	N
P17	Pregunta 17	NUMBER(3)	N
P18	Pregunta 18	NUMBER(3)	N
P19	Pregunta 19	NUMBER(3)	N
P20	Pregunta 20	NUMBER(3)	N
P21	Pregunta 21	NUMBER(3)	N
P22	Pregunta 22	NUMBER(3)	N
P23	Pregunta 23	NUMBER(3)	N
P24	Pregunta 24	NUMBER(3)	N
P25	Pregunta 25	NUMBER(3)	N
P26	Pregunta 26	NUMBER(3)	N
P27	Pregunta 27	NUMBER(3)	N
P28	Pregunta 28	NUMBER(3)	N
P29	Pregunta 29	NUMBER(3)	N
CC1	Calificación de concepto 1	NUMBER(3)	N
CC2	Calificación de concepto 2	NUMBER(3)	N
CC3	Calificación de concepto 3	NUMBER(3)	N
CC4	Calificación de concepto 4	NUMBER(3)	N
CC5	Calificación de concepto 5	NUMBER(3)	N
CC6	Calificación de concepto 6	NUMBER(3)	N

Tabla 7.6 Datos de las preguntas de la Encuesta

7.2 Desarrollo de Minería de Datos

Durante el desarrollo de este apartado se utilizaron herramientas de minería de datos, RapidMiner y WEKA y de estadística PASW Statistics v.18 para el análisis previo de la información, con el objetivo de conocer los aspectos que el profesor necesita mejorar, mediante algunos elementos que se pueden extraer de la encuesta y que le permitan mejorar su proceso de enseñanza-aprendizaje.

A través de RapidMiner y WEKA utilizando matrices de correlación, árboles de decisión, agrupamientos y redes neuronales se logró observar el comportamiento y relación entre los atributos a nivel Facultad, división y departamento y a través de PASW Statistics se logró observar gráficas de tendencias que permiten observar con mayor claridad los patrones obtenidos con anterioridad.

Cabe destacar, que los resultados que arroje el sistema sobre una asignatura están basados en un histórico que data desde 1998, y que contiene la información de las encuestas sobre cada profesor, realizadas semestre con semestre.

Como parte del proceso de difusión y uso del conocimiento extraído se realizó un manual de usuario sobre las herramientas utilizadas para minería de Datos, es decir, WEKA y RapidMiner; en otras palabras, un tutorial con el fin de que se pueda seguir haciendo minería de datos sobre las encuestas de profesores.

7.2.1 Selección de Datos

La selección de datos consiste en el diseño de una vista minable que es una tabla o arreglo de la cual se puede obtener información útil y a la cual se le aplicó minería de datos. La primera fase de nuestra selección de datos fue realizar un proceso de correlación entre las preguntas de la encuesta y seleccionar las que tienen mayor relevancia en el proceso de enseñanza-aprendizaje. En la tabla 7.7 se muestran las correlaciones que se encontraron entre las preguntas de la encuesta sobre el profesor y el alumno.

De las siguientes preguntas se integraron solo las que se consideran válidas respecto al capítulo V.

El profesor	El alumno
Confianza que inspira en los alumnos para intervenir en clase (pregunta 02)	Participo activamente en clase (pregunta24)
Propicia que los alumnos amplíen por su cuenta sus conocimientos sobre la asignatura (pregunta 06)	Me siento motivado para estudiar la asignatura (pregunta 25)
Los conocimientos del profesor sobre la asignatura son (pregunta 05)	Mi aprendizaje de los temas ha sido (pregunta 26)
Al definir conceptos se presentan ejemplos que facilitan su comprensión (pregunta 18)	Realizo tareas y trabajos que me asigna el profesor (pregunta 27)
Adecuado grado de dificultad en exámenes y/o evaluaciones (pregunta 13)	Mi aprendizaje de los temas ha sido (pregunta 26)

Tabla 7.7 Preguntas Alumno- Profesor

En la tabla 7.8 se muestran las correlaciones que se encontraron sobre las preguntas del profesor.

El alumno	El alumno
Falté a clases durante el semestre (pregunta 21)	Mi aprendizaje de los temas ha sido (pregunta 26)
Me siento motivado para estudiar la asignatura (pregunta 25)	La actitud del grupo ayudó a mejorar mi desempeño (pregunta 28)
Para esta asignatura estudio por mi cuenta a la semana (pregunta 23)	Realizo tareas y trabajos que me asigna el profesor (pregunta 27)

Tabla 7.8 Preguntas Alumno-Alumno

Por otra parte, a continuación se presenta una lista en la que se muestran las preguntas que son válidas respecto a la revisión de la encuesta realizada en el capítulo V, en la que algunas preguntas son descartadas debido a la incapacidad del alumno para poder contestarlas.

1. La confianza que el profesor inspira a los alumnos para intervenir en clase. (pregunta 2).
2. El profesor expone la clase en forma clara. (pregunta 4).
3. El profesor propicia que los alumnos amplíen por su cuenta sus conocimientos sobre la asignatura. (pregunta 6).
4. El profesor promueve un ambiente de respeto en clase. (pregunta 8).
5. El profesor regresa los exámenes, tareas, trabajos y proyectos corregidos. (pregunta 12).
6. El profesor entrega oportunamente las calificaciones. (pregunta 15).
7. Al definir principios y conceptos se presentan ejemplos que faciliten la comprensión. (pregunta 18).
8. En caso de tener la oportunidad ¿volvería a tomar clase con el profesor? (pregunta 19).
9. Al inicio del curso se dieron a conocer los objetivos. (pregunta 20).

De las tablas 7.7 y 7.8 se integraron las preguntas que son válidas respecto al capítulo V.

10. La confianza que el profesor inspira a los alumnos para intervenir en clase. (pregunta 2).
11. El profesor expone la clase en forma clara. (pregunta 4).
12. El profesor propicia que los alumnos amplíen por su cuenta sus conocimientos sobre la asignatura. (pregunta 6).
13. El profesor promueve un ambiente de respeto en clase. (pregunta 8).
14. El profesor regresa los exámenes, tareas, trabajos y proyectos corregidos. (pregunta 12).
15. El profesor entrega oportunamente las calificaciones. (pregunta 15).
16. Al definir principios y conceptos se presentan ejemplos que faciliten la comprensión. (pregunta 18).
17. En caso de tener la oportunidad ¿volvería a tomar clase con el profesor? (pregunta 19).
18. Al inicio del curso se dieron a conocer los objetivos. (pregunta 20).

19. Para esta asignatura estudio por mi cuenta(pregunta 23).
20. Participo activamente en clase (pregunta 24).
21. Me siento motivado para estudiar la asignatura (pregunta 25).
22. Mi aprendizaje de los temas ha sido (pregunta 26).
23. Realizo tareas y trabajos que me asigna el profesor (pregunta 27).
24. La actitud del grupo ayudó a mejorar mi desempeño(pregunta 28).

7.2.2 Minería de Datos

Por motivos de confidencialidad de los datos, el estudio se realizó por asignatura y en esta fase se utilizaron, únicamente técnicas descriptivas, es decir, los algoritmos correspondientes a las reglas de asociación y agrupamientos.

Primeramente se realizó una clasificación binaria en relación a la calificación del profesor (CP) donde la calificación del profesor va de 0 a 100, si $CP < 70$, $CP = NO$ y si $CP \geq 70$, $CP = SI$; lo anterior, se realizó con el objeto de obtener un panorama general sobre los datos, y conocer el comportamiento general de los profesores por asignatura.

7.2.2.1 Comportamiento de los Datos

Para realizar las gráficas que ilustren el comportamiento de los datos, se usó el software SPSS v.18, en el cual se pueden crear todo tipo de gráficas importando archivos o realizando consultas desde una base de datos; en la imagen siguiente se muestra la pantalla de inicio del software SPSS v.18, en la cual permite al usuario seleccionar una fuente de datos existente, abrir otro tipo de archivos, ingresar los datos o inicializar el asistente de bases de datos; este mismo permite realizar una conexión entre una base de datos y el software de la misma forma en la que se hace con otras herramientas como WEKA y Rapid Miner. En la figura 7.2 se observa la pantalla de inicio de PASW Statistics.

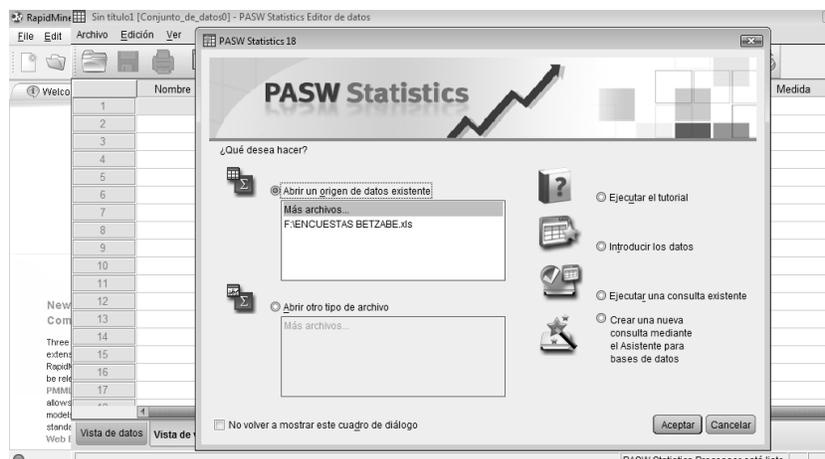


Figura 7.2 Pantalla de Inicio de SPSS v.18

En este caso, se seleccionó un archivo de Excel, una vez abierto el archivo, se pueden realizar tareas de análisis y descriptivas de los datos, una de ellas son los gráficos.

En la figura 7.3 se muestra la opción para seleccionar el generador de gráficos del menú.

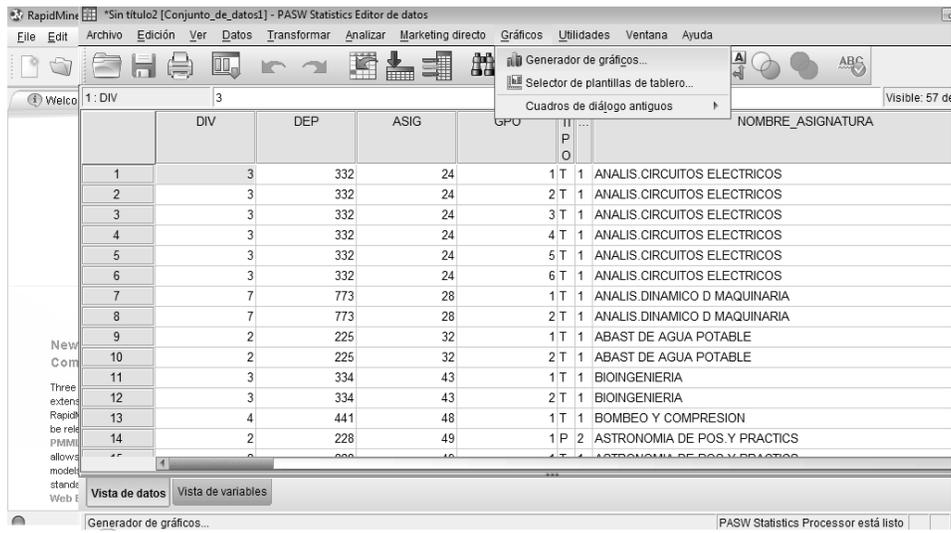


Figura 7.3 Pantalla para seleccionar el generador de gráficos. SPSS v.18

El generador de gráficos tiene la funcionalidad de seleccionar las variables sobre las cuales se desea realizar un gráfico y seleccionar el tipo de función estadística que se desea utilizar para graficar. En la figura 7.4 se muestran algunas de las funciones del generador de gráficos.

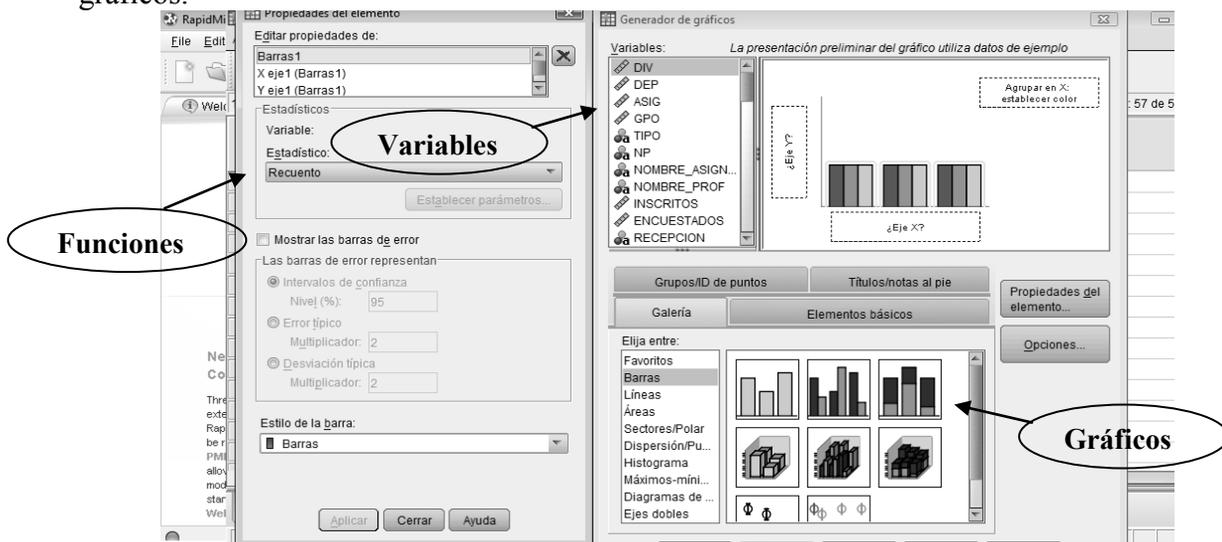


Figura 7.4 Pantalla del generador de gráficos. SPSS v.18

7.2.2.2 Comportamiento de los Profesores

En la figura 7.5 se ilustra el comportamiento de las calificaciones del profesor en relación con el semestre impartido; en el eje de las abscisas se muestra el semestre impartido, que comienza desde el semestre 1998-1 hasta el semestre 2010-1, mientras que en eje de las ordenadas se muestra el recuento de los profesores por semestre y la variable que colorea las barras muestra si el profesor obtuvo una calificación mayor o menor a 70 según sea el caso.

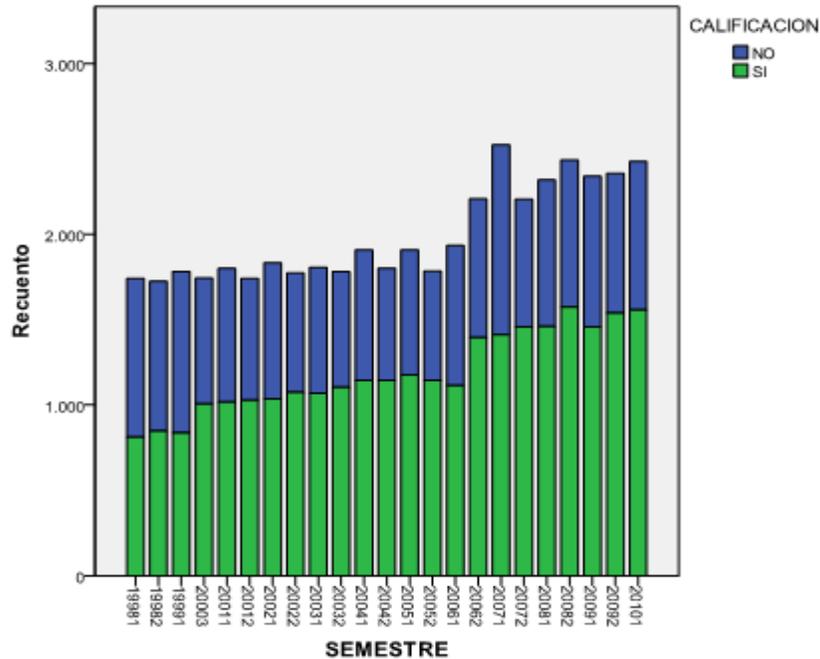


Figura 7.5 Gráfica de calificación del profesor y semestre impartido

En la figura 7.6 se puede observar el comportamiento de los profesores de la Facultad de Ingeniería en cuanto a las tendencias de las calificaciones de los profesores y sus divisiones correspondientes; para ello se muestra en el eje x, la división correspondiente, mientras que en el eje y se muestra el recuento de profesores y su calificación; nuevamente la variable que colorea las barras, muestra si el profesor obtuvo una calificación mayor o menor a 70 y se distinguen con su color correspondiente.

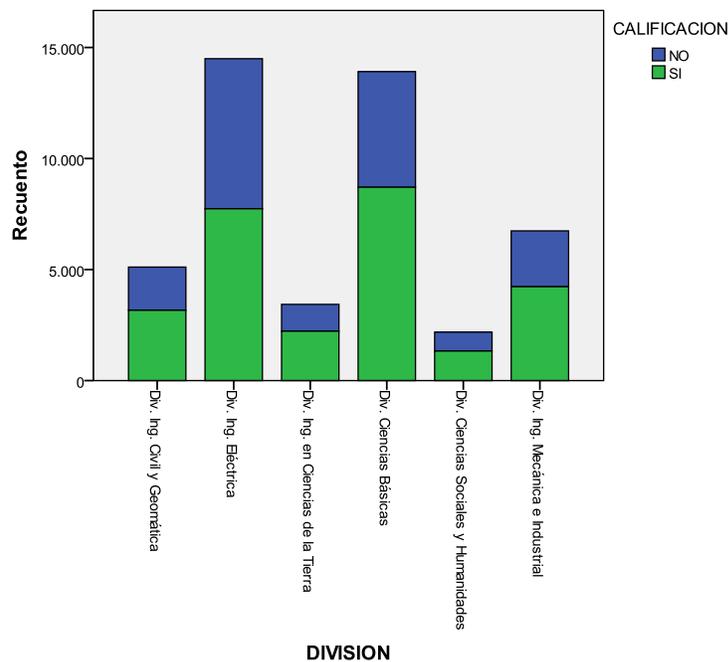


Figura 7.6 Gráfica de Calificación de Profesores por División

Como se puede observar en la figura 7.6, las divisiones con mayor número de profesores es la 2 y la 4, la División de Ingeniería Eléctrica y la División de Ciencias Básicas según correspondan; además se puede observar que la división con mayor número de profesores con calificación general menor a 70 es la División de Ingeniería Eléctrica, dando un total de 45,875 registros.

En la figura 7.7 se muestra en el eje x la calificación de los alumnos, mientras que en el eje y se muestra el recuento de los grupos con calificaciones similares, mientras que la variable que los colorea es si la calificación del profesor es mayor o menor a 70.

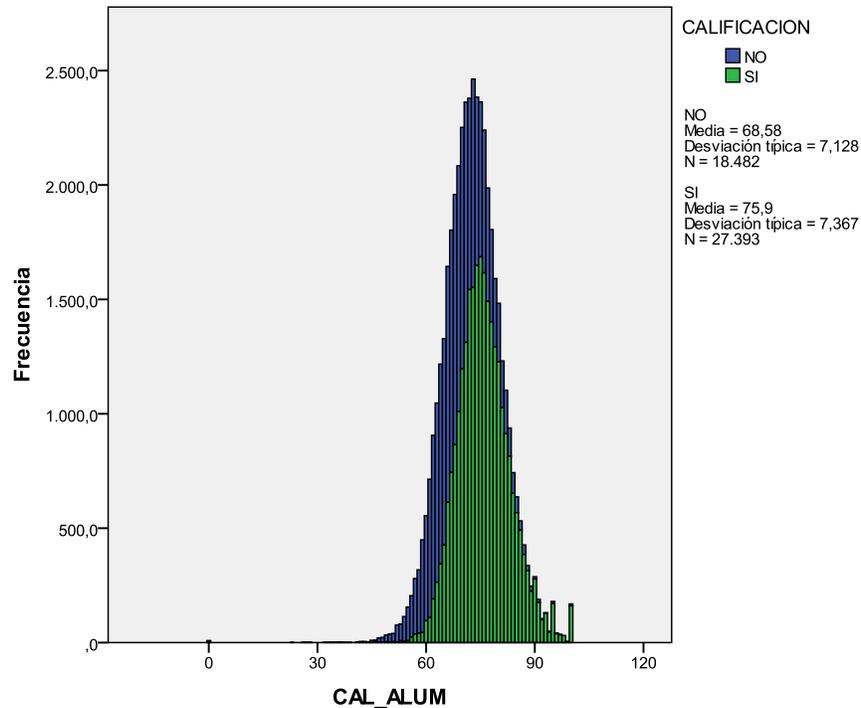


Figura 7.7 Gráfica de Calificación de alumnos en relación con la relación de profesores

Como se puede observar en la figura 7.7, las calificaciones de los alumnos se concentran entre los 66 y los 80 puntos; también se puede observar que mientras es mayor la calificación del alumno, existe una mayor probabilidad de que la calificación del profesor se encuentre por encima de los 70 puntos; y mientras más baja sea la calificación del alumno, existe mayor probabilidad de que la calificación del profesor sea menor a los 70 puntos; este mismo gráfico nos arroja algunos datos estadísticos como la media, la desviación típica y el número de muestras N, tanto para “SI” como para “NO”. La media se refiere al promedio de calificaciones para “SI” y para “NO”; la desviación típica se refiere a la medida de dispersión de los datos, tanto para “SI” como para “NO”.

Debido a los resultados anteriores, se procedió a realizar el mismo análisis pero con una calificación mayor a 80 para “Si” y menor a 80 para “No”, ya que la mayoría de los profesores se encuentra en una calificación general promedio de 70. En la figura 7.8 se puede observar la gráfica de comportamiento de calificación de los alumnos en relación con

la calificación de profesores cuando los profesores tienen una calificación mayor 80 para “SI” y menor a 80 para “NO”.

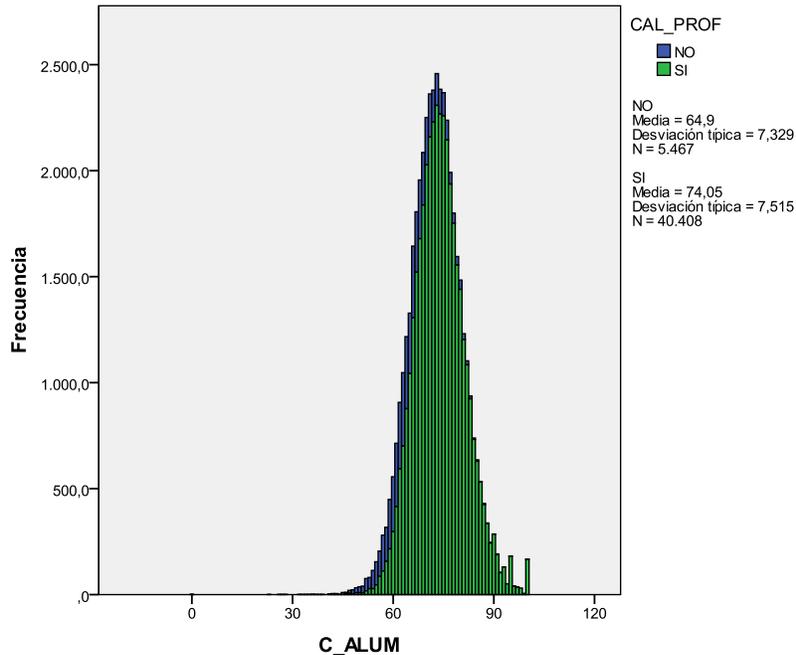


Figura 7.8 Gráfica de Calificación de alumnos en relación con la relación de profesores

En la figura 7.8 se puede observar la media, desviación típica y número de muestras N para cada caso. Se observa que el número de casos para SI aumenta y por consecuente disminuyó para NO, así también disminuyó la desviación típica por lo que se concluye que es un mejor modelo.

7.2.2.3 Matriz de Correlaciones

Para obtener una vista de cómo se encuentran relacionadas las variables entre sí y de esta forma poder seleccionar las más significativas, se realizó a continuación una matriz de correlaciones mediante el software libre Rapid Miner v.5. Para comenzar a utilizar Rapid Miner v.5, debe seleccionarse un archivo nuevo como se muestra en la figura 7.9.

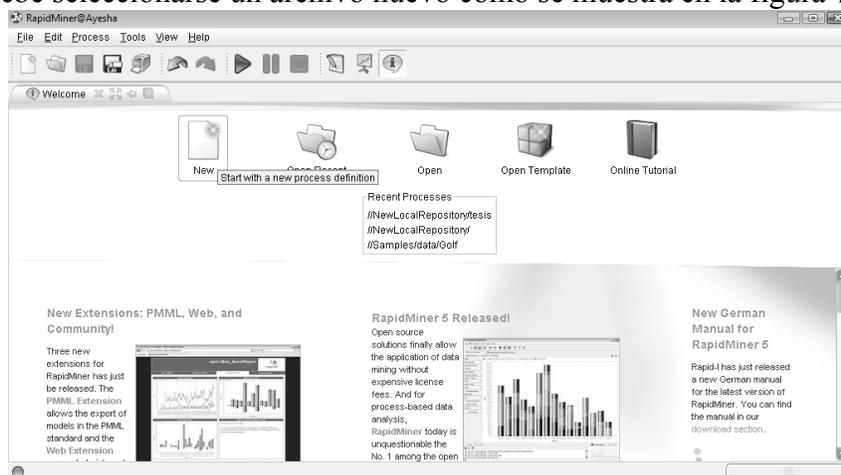


Figura 7.9 Imagen de Inicio de Rapid Miner

Una vez realizado lo anterior, se debe seleccionar el nombre del archivo que tendrá el repositorio de datos. Para ilustrar el paso anterior se muestra la figura 7.10.

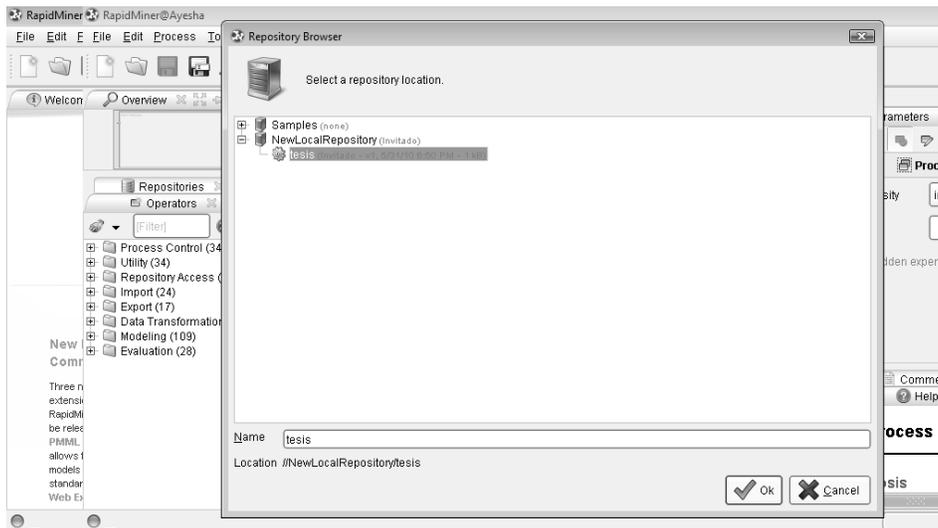


Figura 7.10 Imagen de Selección del Repositorio

Una vez seleccionado el repositorio, Rapid Miner abre un archivo nuevo en el cual deben de ser importados los datos, en este caso se importó un archivo de Excel. Esta opción se encuentra en la opción Import, se selecciona Data y posteriormente se selecciona Read Excel; para concluir con este paso, la opción seleccionada se arrastra hasta el área dispuesta para bloques como se muestra en la figura siguiente.

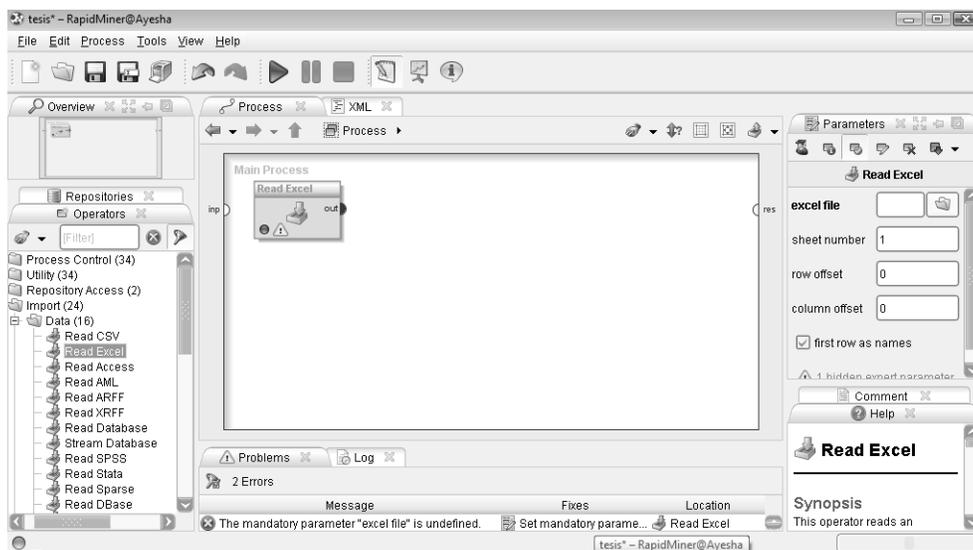


Figura 7.11 Pantalla de selección de Importar archivo de Excel

Para que el bloque anterior sea habilitado, debe seleccionarse un archivo a importar; esto se puede realizar seleccionando el archivo en el bloque superior derecho en la pantalla anterior.

En la figura 7.12 se muestra el archivo de Excel importado.

	A	B	C	D	E	F	G	H	M	N	BA	BB	BC	BD	BE
1	LABEL	DIV	DEP	ASIG	GPO	TIPO	NP	NOMBRE_ASIGNATURA	CAL_PROF	CAL_ALUM	CALIFICACION	ASIG	C_ALUM	ASIG	
2	1	3	332	24	1	T	1	ANALIS.CIRCUITOS ELECTRICOS	80	71	SI	24	71	24	
3	2	3	332	24	2	T	1	ANALIS.CIRCUITOS ELECTRICOS	83	71	SI	24	71	24	
4	3	3	332	24	3	T	1	ANALIS.CIRCUITOS ELECTRICOS	85	77	SI	24	77	24	
5	4	3	332	24	4	T	1	ANALIS.CIRCUITOS ELECTRICOS	71	65	SI	24	65	24	
6	5	3	332	24	5	T	1	ANALIS.CIRCUITOS ELECTRICOS	79	73	SI	24	73	24	
7	6	3	332	24	6	T	1	ANALIS.CIRCUITOS ELECTRICOS	91	79	SI	24	79	24	
8	7	7	773	28	1	T	1	ANALIS.DINAMICO D MAQUINARIA	69	60	NO	28	60	28	
9	8	7	773	28	2	T	1	ANALIS.DINAMICO D MAQUINARIA	64	59	NO	28	59	28	
10	9	2	225	32	1	T	1	ABAST DE AGUA POTABLE	77	63	SI	32	63	32	
11	10	2	225	32	2	T	1	ABAST DE AGUA POTABLE	66	64	NO	32	64	32	
12	11	3	334	43	1	T	1	BIOINGENIERIA	78	66	SI	43	66	43	
13	12	3	334	43	2	T	1	BIOINGENIERIA	66	65	NO	43	65	43	
14	13	4	441	48	1	T	1	BOMBEO Y COMPRESION	75	68	SI	48	68	48	
15	14	2	228	49	1	P	2	ASTRONOMIA DE POS. Y PRACTICS	88	83	SI	49	83	49	
16	15	2	228	49	1	T	1	ASTRONOMIA DE POS. Y PRACTICS	77	67	SI	49	67	49	
17	16	5	553	56	2	T	1	FISICA EXPERIMENTAL (GEN 94)	83	71	SI	56	71	56	
18	17	5	553	56	2	P	2	FISICA EXPERIMENTAL (GEN 94)	67	71	NO	56	71	56	
19	18	5	553	56	3	P	2	FISICA EXPERIMENTAL (GEN 94)	76	70	SI	56	70	56	
20	19	5	553	56	3	T	1	FISICA EXPERIMENTAL (GEN 94)	85	75	SI	56	75	56	
21	20	5	553	56	2	T	1	FISICA EXPERIMENTAL (GEN 94)	86	70	SI	56	70	56	

Figura 7.12 Archivo de Excel para matriz de correlaciones

Si lo que se desea es realizar una matriz de correlaciones, debe de seleccionarse del menú desplegable la opción Modeling, a continuación Correlation and Dependency Computation y por último la opción Correlation Matrix; igual que en el procedimiento anterior, debe de arrastrarse hasta el área de trabajo la función seleccionada. Por último, ambos bloques deben unirse dando clic sobre la salida del bloque y sin soltar el botón, arrastrar hasta la entrada del siguiente bloque, una vez ubicado el puntero del mouse donde se quiere conectar se suelta el botón; de la misma forma se conectan sus salidas hacia los archivos que se van a crear. En la figura 7.15 se muestra el diagrama de bloques completo para la obtención de una matriz de correlaciones.

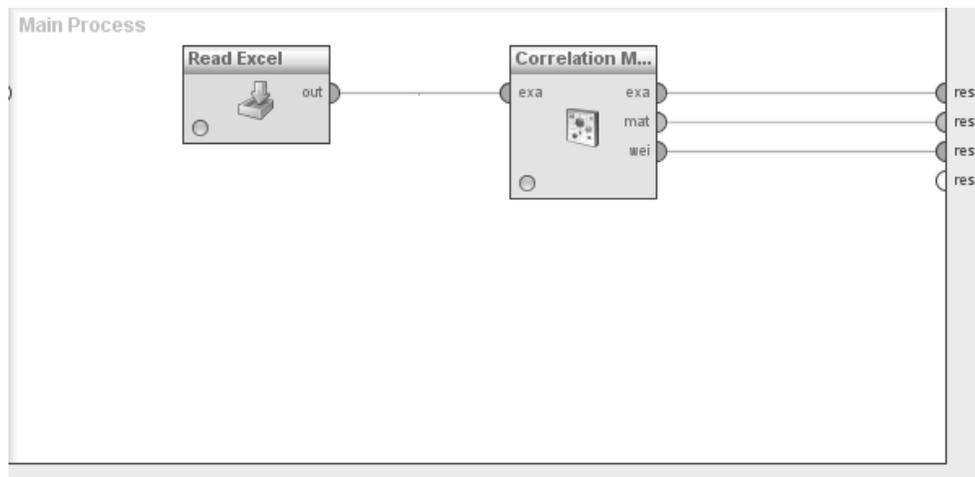


Figura 7.13 Diagrama de bloques en Rapid Miner

A continuación Rapid Miner muestra el resultado de la ejecución del diagrama de bloques, y muestra la Matriz de Correlaciones obtenida.

En la figura 7.14, se muestra la pantalla de RapidMiner después de su ejecución.

Figura 7.14 Matriz de Correlaciones en Rapid Miner

De la Matriz de Correlación obtenida en el paso anterior, se pueden destacar algunas variables que tienen un valor de correlación muy alto; Rapid Miner realiza lo anterior resaltando los valores altos con celdas de diferente color. Cabe señalar que de la Matriz de correlaciones obtenida en el paso anterior, se debe realizar una depuración de resultados, ya que algunas correlaciones muy altas pueden deberse a que es la misma pregunta o no tienen coherencia; en la siguiente tabla se muestran los valores de las correlaciones más altas entre los atributos. En la tabla 7.9 se muestran las correlaciones más altas del análisis anterior.

Atributo	Atributo	Porcentaje de Correlación
CAL_ALUM	P28	0.8044733977541743
CAL_PROF	P18	0.8050980085377885
P03	P04	0.8056772529617363
P04	P18	0.8108055364002485
CAL_PROF	P19	0.8115983533660752
P28	C_ALUM	0.8128584526148508
P18	C_PROF	0.8132478306695848
P03	C_PROF	0.8159991693452691
P18	CC2	0.8171791594030534
P04	P19	0.818229046473337
P19	C_PROF	0.8198150935097078
CAL_PROF	P07	0.8220821581276287
P10	CC2	0.8224023875703693
P10	P18	0.8229604825041945
P18	CC1	0.8249827211845836
P01	CC1	0.8253736171132658
CAL_PROF	P04	0.8287049252891269
P04	CC1	0.8287102521973527
CAL_PROF	P10	0.8306635536912709
P07	C_PROF	0.8307441511071
P04	C_PROF	0.8364622396739112
P05	CC1	0.8385519548550121
P10	C_PROF	0.8386603976063058
P07	CC2	0.8409484504489348
CAL_PROF	P01	0.8422121014737518
P09	CC1	0.8472872618338217

P01	C_PROF	0.850359539989505
P19	CC2	0.8595035386278422
P21	CC6	0.8669890795921967
P02	CC2	0.8710024426320662
P12	P15	0.8841855964514854
CC1	CC2	0.8844840309879846
P10	CC1	0.8932561662898051
P07	CC1	0.8953164905730568
CAL_PROF	CC2	0.8977260353122162
P04	CC2	0.8980870632268501
CAL_PROF	CC1	0.8984156650414851
P01	CC2	0.9028301075707161
P03	CC2	0.9082341077273526
P15	CC3	0.9562062036020030
P12	CC3	0.9577540950320129

Tabla 7.9 Correlaciones más altas

Se puede realizar un ordenamiento de datos, dando clic sobre la variable que queremos ordenar. Una vez analizada la relación entre los atributos, se pudo observar que la correlación entre las preguntas seleccionadas en el apartado anterior es alta y mostró algunas relaciones que no se habían tomado en cuenta, debido a esto, fueron agregadas algunas correlaciones que no habían sido consideradas; por lo que se procedió a la construcción de árboles de decisión que ilustren de una mejor forma la relación de estos atributos con las demás columnas. Debido a los resultados anteriores, se integraron algunas correspondencias a la tabla de selección de datos como se muestra en la tabla 7.10.

El profesor	El alumno
Confianza que inspira en los alumnos para intervenir en clase (pregunta 02)	Participo activamente en clase (pregunta24)
Propicia que los alumnos amplíen por su cuenta sus conocimientos sobre la asignatura (pregunta 06)	Me siento motivado para estudiar la asignatura (pregunta 25)
Los conocimientos del profesor sobre la asignatura son (pregunta 05)	Mi aprendizaje de los temas ha sido (pregunta 26)
Al definir conceptos se presentan ejemplos que facilitan su comprensión (pregunta 18)	Realizo tareas y trabajos que me asigna el profesor (pregunta 27)
Adecuado grado de dificultad en exámenes y/o evaluaciones (pregunta 13)	Mi aprendizaje de los temas ha sido (pregunta 26)
Calificación de Concepto de Motivación	Calificación General del Profesor
Las clases se desarrollan de manera Interesante(pregunta 3)	Calificación de Concepto de Motivación
Entrega oportunamente las calificaciones(pregunta 15)	Calificación de Concepto de Evaluación
Regresa los exámenes, tareas, trabajos y proyectos (pregunta 12)	Calificación de Concepto de Evaluación
Calificación de Concepto de Aprendizaje	Calificación General del Profesor
Agregadas	
Calificación del Alumno	La actitud del grupo ayudó a mejorar mi desempeño (Pregunta 28)
Calificación del Profesor	Se presentan ejemplos que facilitan la comprensión (Pregunta 18)
El profesor expone la clase en forma clara (pregunta 4)	Se presentan ejemplos que facilitan la comprensión (Pregunta 18)
Calificación del Profesor	En caso de tener oportunidad ¿volvería a tomar clases con el profesor?(Pregunta 19)
Fomenta la habilidad de plantear y resolver problemas(Pregunta 10)	Motivación(CC2)
Se presentan ejemplos que facilitan la comprensión	Aprendizaje(CC1)

(Pregunta 18)	
El interés del profesor por que los alumnos aprendan es(pregunta 1)	Aprendizaje(CC1)
El interés del profesor por que los alumnos aprendan es(pregunta 1)	Motivación(CC2)
Las clases se desarrollan de manera (pregunta 3)	Motivación(CC2)
El profesor entrega oportunamente las calificaciones(Pregunta 15)	Evaluación(CC3)
El profesor regresa los exámenes, tareas, trabajos y proyectos corregidos (pregunta 12)	Evaluación(CC3)

Tabla 7.10 Correlación de preguntas válidas

7.2.2.4 Árboles de Decisión

Mediante Rapid Miner se puede realizar la construcción de árboles de decisión; cabe destacar, que los árboles de decisión, corresponden tanto a tareas predictivas, como descriptivas de Minería de Datos. Como primer paso se debe de seleccionar una fuente de datos, en este caso se selecciona la opción de importar datos de Excel del menú desplegable en la opción “Data”. Posteriormente se selecciona del menú, en la opción de modelado, en el apartado de “Tree induction” el algoritmo correspondiente, en este procedimiento se seleccionó el algoritmo Random Forest; a continuación, Rapid Miner solicita que sea seleccionado del archivo de Excel, una variable que tenga la funcionalidad de “label”, es decir, la variable sobre la que se quiere obtener patrones, y una variable que sirva como llave primaria, en este caso se seleccionó la variable Calificación del Profesor. Una vez seleccionadas estas opciones, se debe de proceder a seleccionar cuántos árboles deseamos que realice, durante esta ejecución le pedimos que realizara 20 árboles. Finalmente para verificar que todo esté conectado y funcionando correctamente, debemos de verificar que se encuentren en color verde los botones de cada uno de los bloques, y que en la parte inferior de la ventana de Rapid Miner, se encuentre la leyenda “No Errors”. En la figura 7.15 se muestra el diagrama de bloques para la construcción de árboles de decisión.

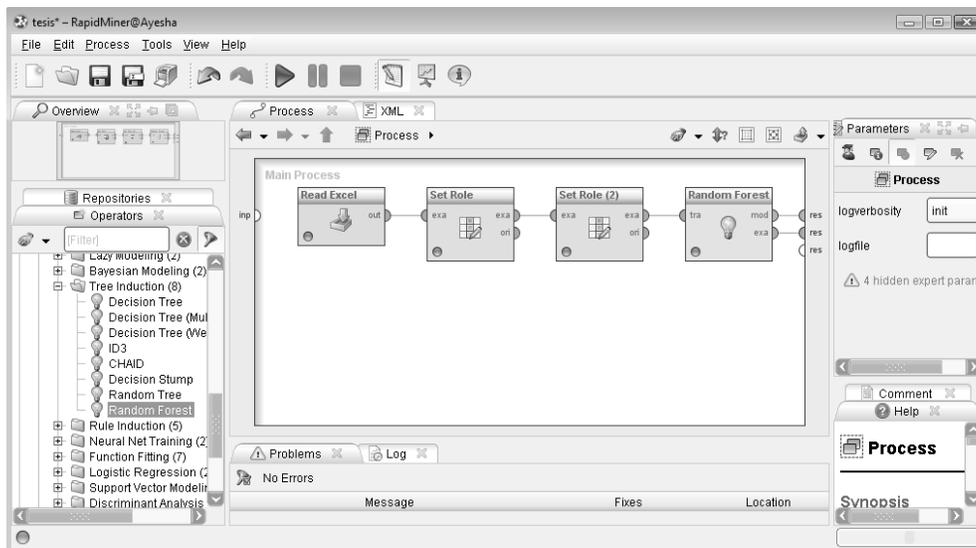


Figura 7.15 Diagrama de Bloques de árboles de Decisión en Rapid Miner

En la figura 7.16 se muestra el archivo de Excel utilizado para árboles de decisión.

	A	B	C	D	E	F	G	H	I	J	K	L
1	DIV	TIPO	CAL_PROF	CAL_ALUM	CALIFICACION	P01	P02	P03	P04	P05	P06	P07
13735	3	L	89	78	NO	100	100	78	89	89	89	100
13736	3	L	84	67	NO	80	90	83	87	83	73	83
13737	3	L	93	90	SI	100	100	83	100	83	83	100
13738	3	L	89	87	SI	89	89	89	93	89	93	96
13739	3	L	75	82	SI	80	80	77	83	77	73	77
13740	3	L	91	87	SI	100	100	97	93	90	90	97
13741	3	L	73	82	SI	83	80	70	90	80	77	80
13742	3	L	83	81	SI	87	97	80	87	90	77	83
13743	3	L	77	76	NO	78	81	78	85	78	70	78
13744	3	L	86	80	NO	87	90	90	93	87	83	87
13745	3	L	70	62	NO	73	73	67	67	67	53	67
13746	3	L	75	68	NO	63	63	57	57	70	63	63
13747	3	L	74	73	NO	78	81	70	81	74	85	81
13748	3	L	70	72	NO	73	73	60	80	73	60	67
13749	3	L	67	74	NO	73	73	67	73	73	60	67
13750	3	L	81	85	SI	85	89	70	81	93	81	81

Figura 7.16 Archivo de Excel Árboles de Decisión

En los árboles desarrollados, se seleccionó la opción “gini index” ya que prueba todos los parámetros existentes en los datos, se hicieron 20 árboles de los cuales, solo algunos son los más significativos. En la figura 7.17, se muestra la pantalla de resultados de los árboles en Rapid Miner.

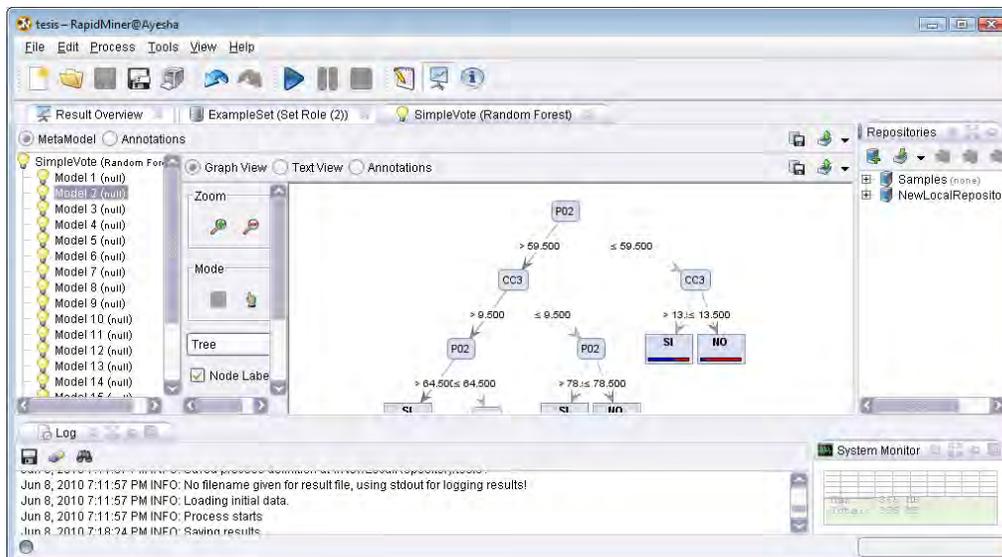
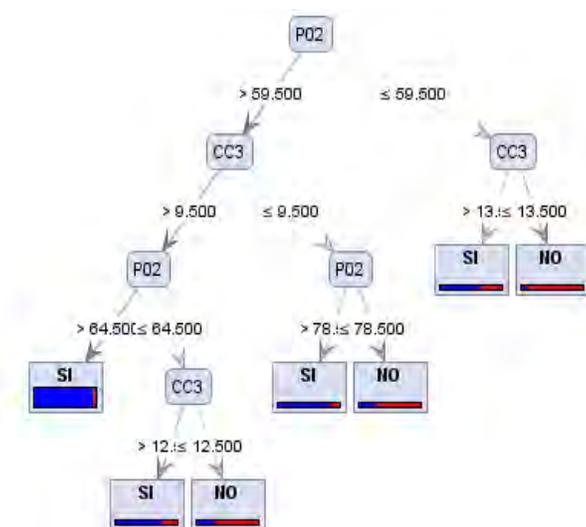


Figura 7.17 Árboles de Decisión Rapid Miner

A continuación se mostrarán algunos de los árboles obtenidos; dichos árboles fueron obtenidos sobre la variable de “Calificación” la cual asigna un valor de “SI” en el caso de que el profesor haya obtenido una calificación mayor a 80, y asigna un valor de “NO” en el caso de que el profesor haya obtenido una calificación menor a 80, esto se puede observar

en las hojas⁹ del árbol; sin embargo las ramas de cada árbol se harán respecto a la calificación de la pregunta del nodo que puede ir del 0 al 100.

El primer árbol obtenido muestra la relación que existe entre la variable P02 y la variable CC3; en la primera rama se encuentra la mayoría de los datos, en este caso son 43424 registros, en los que cuando el valor de P02 es mayor a 59.500 de calificación se evalúa CC3 y si este nodo tiene de calificación mayor a 9.5 y P02 es mayor a 64.5 muestra que la mayoría de los profesores tienen calificación mayor a 80, es decir obtienen “SI”.



Nodo	Descripción
CC3	Calificación de concepto de Evaluación
P02	La confianza que el profesor inspira a los alumnos es

Árbol de Decisión 1. Confianza que el profesor inspira a sus alumnos

Cabe señalar, que para conocer el tamaño y la frecuencia de cada nodo, solo es necesario situarse con el puntero del mouse sobre el nodo del cual se requiere conocer sus dimensiones, tal y como se muestra en la figura 7.18.

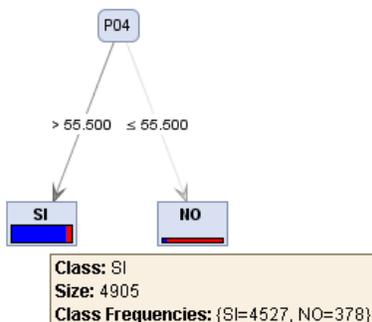
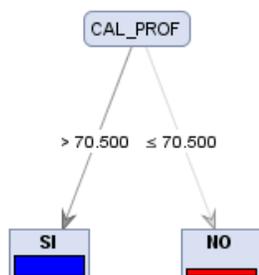


Figura 7.18 Frecuencia de los nodos

⁹ Nodos que no tienen ramas o conexión a otros nodos, también conocidos como nodos terminales.

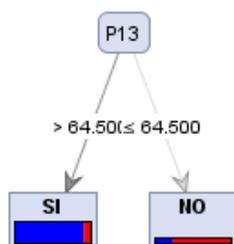
En el árbol de decisión 2 se hace un agrupamiento sobre la calificación del profesor, en este caso, la mayor población de profesores obtienen una calificación mayor a 80. En este caso podemos observar que cuando la calificación del nodo CAL_PROF es mayor a 70.5, los profesores obtienen una calificación mayor a 80, es decir que obtienen “SI”.



Nodo	Descripción
CAL_PROF	Calificación del Profesor

Árbol de Decisión 2. Calificación del profesor

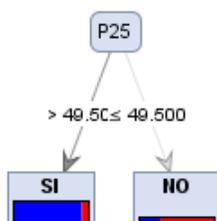
En el árbol de decisión 3 se realizó un agrupamiento de la calificación de profesores respecto a la pregunta 13 que se refiere al adecuado grado de dificultad en exámenes y/o evaluaciones, en la que 43236 registros, obtienen una calificación mayor a 64.500 de 45875 registros totales.



Nodo	Descripción
P13	Los exámenes y/o evaluaciones tienen un grado de dificultad adecuado

Árbol de Decisión 3. Grado Adecuado de las evaluaciones del profesor

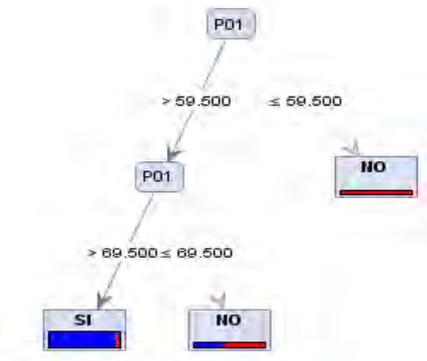
En el árbol de decisión 4, se realiza un agrupamiento de registros sobre la variable de calificación del profesor, en base a la calificación obtenida en la pregunta P25 que se refiere a la motivación para estudiar la asignatura. En la siguiente figura se observa que la mayoría de los registros presenta un comportamiento en el cual, obtienen una calificación menor a 49 en esa pregunta, sin embargo, los profesores obtienen una calificación mayor a 80 en esas condiciones.



Nodo	Descripción
P25	Me siento motivado para estudiar la asignatura

Árbol de Decisión 4. Motivación para estudiar la asignatura

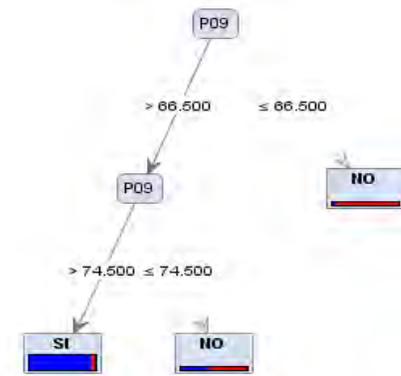
En el árbol de decisión 5 se ilustra los agrupamientos realizados sobre la información en relación a la pregunta 1, la cual hace referencia al interés por parte del profesor por que los alumnos aprendan.



Nodo	Descripción
P01	El interés del profesor por que los alumnos aprendan es

Árbol de Decisión 5. Interés del profesor

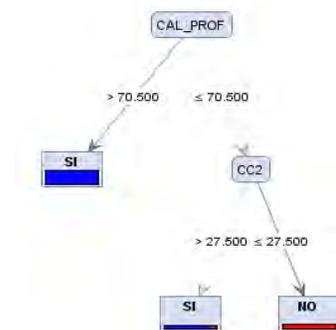
En el árbol de decisión 6 se ilustra los agrupamientos realizados sobre la información en relación a la pregunta 9, la cual hace referencia a si las tareas, trabajos y/o ejercicios contribuyen al aprendizaje.



Nodo	Descripción
P09	Las tareas, trabajos y/o ejercicios que deja el profesor contribuyen al aprendizaje

Árbol de Decisión 6. Tareas y trabajos que deja el profesor

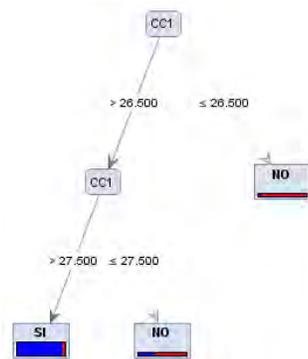
En el árbol de decisión 7 se ilustra los agrupamientos realizados sobre la información en relación a la calificación del profesor y su relación respecto a la calificación de concepto 2 que se refiere a Motivación. En la figura siguiente se puede observar, que los profesores que obtienen una calificación general menor a 70.500, tienen una calificación de concepto respecto a motivación baja.



Nodo	Descripción
CAL_PROF	La confianza que el profesor inspira a los alumnos es
CC2	Calificación de Concepto de Motivación

Árbol de Decisión 7. Calificación del profesor en relación con la motivación

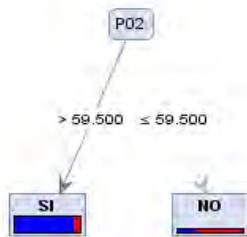
En el árbol de decisión 8 se ilustra los agrupamientos realizados sobre la información en relación al concepto CC1 que evalúa el Aprendizaje.



Nodo	Descripción
CC1	Calificación de Concepto de Aprendizaje

Árbol de Decisión 8. Aprendizaje

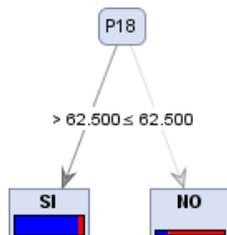
En el árbol de decisión 9 se ilustra los agrupamientos realizados sobre la información en relación a la pregunta 2, la cual hace referencia a si el profesor inspira confianza a los alumnos para intervenir en clase.



Nodo	Descripción
P02	La confianza que el profesor inspira a los alumnos es

Árbol de Decisión 9. Confianza que inspira a los alumnos

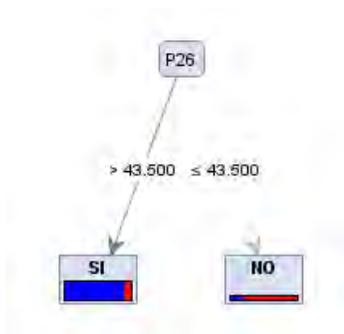
En el árbol de decisión 10 se ilustra los agrupamientos realizados sobre la información en relación a la pregunta 18, la cual hace referencia a si el profesor presenta ejemplos que facilitan la comprensión de los temas.



Nodo	Descripción
P18	Se presentan ejemplos que facilitan la comprensión

Árbol de Decisión 10. Ejemplos que faciliten la comprensión

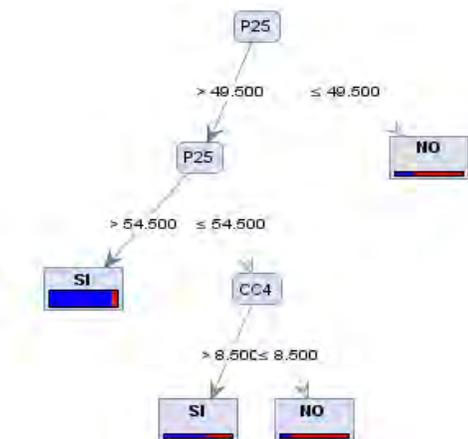
En el árbol de decisión 11 se ilustra los agrupamientos realizados sobre la información en relación a la pregunta 26, la cual hace referencia a si el aprendizaje sobre los temas del alumno es bueno.



Nodo	Descripción
P26	Mi aprendizaje sobre los temas ha sido

Árbol de Decisión 11. Aprendizaje de los temas

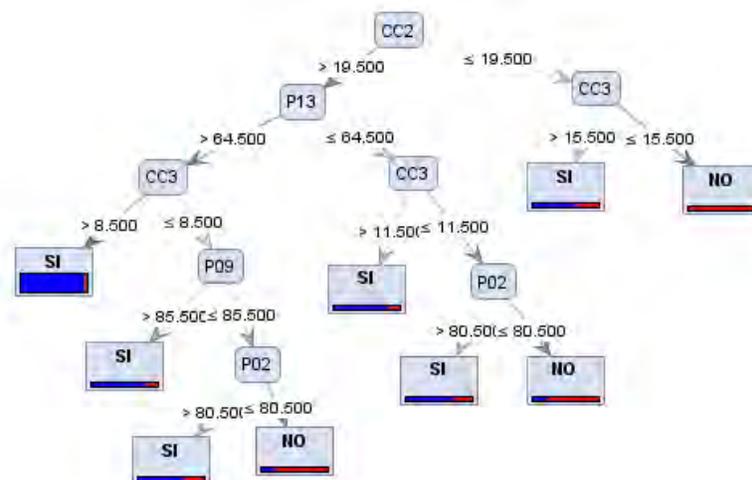
El primer árbol obtenido muestra la relación que existe entre la variable P25(Si el alumno se encuentra motivado para estudiar la asignatura) y la variable CC4 (Responsabilidad); en la primera rama se encuentra la mayoría de los datos, en los que el valor de P25 supera el 49.500 de calificación, y en el cual la mayor cantidad de datos se sitúa donde CC4 tiene un puntaje mayor a 8.500; siguiendo la misma rama, observamos que para esta variante, P25 tiene la mayor cantidad de registros situados en una calificación mayor a 54.500.



Nodo	Descripción
P25	Motivación para estudiar la asignatura
CC4	Calificación de Concepto de Responsabilidad

Árbol de Decisión 12. Responsabilidad y Motivación

A continuación se presenta el árbol más frondoso obtenido durante este procedimiento, en el cual, se presentan algunos de los atributos más significativos y su relación correspondiente. Lo primero que se puede observar es que la mayor cantidad de las encuestas realizadas obtienen un puntaje mayor a 19.500 en su calificación del concepto de Motivación, como un subgrupo se presenta a la pregunta 3, la cual hace referencia a si las clases se desarrollan de manera interesante y como se observa, la mayor cantidad de registros obtienen una calificación mayor a 64.500 en esa pregunta; a su vez, esta pregunta se clasifica en dos ramas haciendo referencia la calificación de concepto de Evaluación.



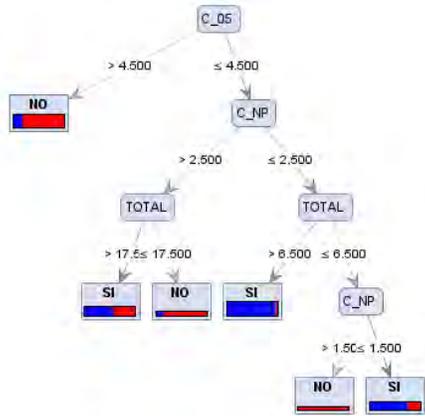
Árbol de Decisión 13. Evaluación, Confianza y Motivación

Nodo	Descripción
CC3	Calificación de Concepto de Evaluación
P02	La confianza que el profesor inspira a los alumnos es
CC2	Calificación de Concepto de Motivación
P09	Las tareas, trabajos y/o ejercicios que deja el profesor contribuyen a mi aprendizaje
P13	Los exámenes y evaluaciones tienen un grado de dificultad adecuado

7.2.2.4.1 Árboles de Decisión por división

De los procedimientos realizados hasta el momento en este capítulo, se ha observado que existen algunas preguntas y algunos elementos de la encuesta que influyen de forma significativa en la calificación del profesor; sin embargo, es necesario que se realice un análisis sobre la influencia que existe entre estos elementos y el desempeño de los alumnos, por lo que se hará uso de la base de datos que contiene el número de aprobados, número de reprobados y número de NP por grupo, de tal forma, que se analizarán las preguntas y conceptos de la encuesta y la cantidad de alumnos reprobados y aprobados, así como los alumnos que no presentaron la asignatura.

En el árbol de decisión 14 se muestra el árbol con más ramificación obtenido después de haber ejecutado el algoritmo Random Forest de árboles de decisión sobre la tabla de datos que incluye el número de aprobados, número de reprobados y número de NP correspondiente a cada profesor, así como las preguntas válidas descritas en el capítulo 5 de este trabajo.



Nodo	Descripción
C_05	Número de alumnos por grupo con calificación de 5
C_NP	Número de alumnos por grupo con NP
TOTAL	Número total de alumnos por grupo

Árbol de Decisión 14. con calificaciones de alumnos

En el árbol anterior se muestran los porcentajes de alumnos con 5 y aprobados, así como con NP que tiene un profesor cuya calificación es mayor (SI) o menor (NO) de 80 puntos. La siguiente tabla es obtenida después de ejecutar el algoritmo K-Means de agrupamiento, sobre los datos anteriormente mencionados; se han obtenido 6 grupos de datos cuyas características principales se muestran en la tabla 7.11.

Atributo	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5	Clúster 6
APROBADOS	17.2464	18.0772	16.3109	13.5997	20.2727	19.1156
C_05	3.00893	4.3057	2.8377	1.76199	3.4569	6.7365
C_NP	4.3117	4.9843	3.6485	2.4477	4.6366	6.9606
TOTAL	24.5671	27.3673	22.7972	17.8095	28.3663	32.8132
%APROBADOS	72.5024	67.3987	72.6693	78.1066	71.7156	60.3750
%REPROBADOS	9.7325	13.3484	10.7813	7.4394	10.4854	17.0997
P02	82.3813	79.7755	82.5264	80.2723	79.1650	78.9867
P04	81.5574	80.1522	80.703	78.6098	77.6495	78.9924
P06	78.1146	76.58027	78.13908	77.16840	76.36337	75.91392
P08	90.46737	89.0544	89.7701	89.3227	88.4676	88.8810
P12	73.4508	75.8351	71.8446	73.1686	72.1503	75.8159
P15	74.0327	76.0698	73.3910	73.8917	72.6188	76.2702
P18	81.9129	83.1344	81.8789	82.5816	81.5049	82.9154
P19	78.0628	76.6883	77.9034	75.5394	73.8024	75.5136
P20	95.5117	95.8312	95.3585	95.5280	95.1930	95.6487
CC1	32.1432	32.3774	31.9777	32.2187	31.9343	32.2807
CC2	25.6723	25.2111	25.5957	25.0264	24.7821	24.9422
CC3	13.6714	13.9701	13.5036	13.5851	13.3871	13.9703
CC4	10.5547	10.2099	10.5406	10.1927	10.1406	10.4663
CC5	47.4012	44.6216	48.1905	44.2281	43.8803	43.2409
CC6	25.3526	23.1380	24.9660	22.7621	22.9352	23.2326
C_PROF	81.6776	81.7509	82.0571	82.4157	80.9596	79.4360
C_ALUM	74.1541	72.8830	72.9094	73.7880	71.7803	70.8105

Tabla 7.11 Resultados de Agrupamiento

De la tabla 7.11 se puede observar, que en la Facultad de Ingeniería el porcentaje promedio de alumnos aprobados por grupo es 70.4612%. De la misma forma se puede observar que el grupo con mayor cantidad de alumnos reprobados es el grupo 6 cuyo número de alumnos es menor al de los demás y su porcentaje de aprobados es menor al de los otros grupos.

A continuación se observará el comportamiento de los datos por división, de la misma forma en que se hizo un análisis general de los datos, mediante el software SPSS v18. En la figura 7.18 se puede observar las divisiones que conforman la Facultad de Ingeniería en el eje “x”, mientras que en el eje “y” se muestra el recuento de datos por división que corresponden a la cantidad de aprobados por grupo, tal que si en un grupo más del 70% aprobó, se obtiene un “SI” y si menos del 70% aprobó se obtiene un “NO”, y de esta forma son coloreadas las barras de la gráfica; se toma este criterio ya que en los agrupamientos anteriores se observa que los porcentajes de aprobados por grupo son en promedio del 70% y si se tomara un porcentaje mayor como referencia, implicaría observar en la gráfica una mayor cantidad de color correspondiente a “NO” por barra.

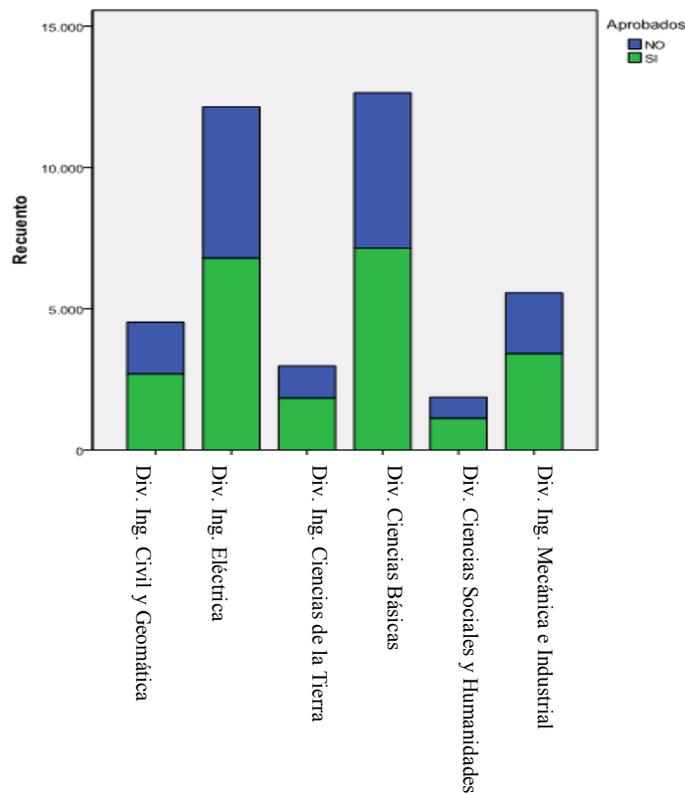


Figura 7.18 Gráfica de Desempeño por división

En la figura 7.19 se muestra la gráfica de la media de alumnos aprobados por división, en la que se puede observar, que la división con mayor porcentaje de alumnos aprobados por curso es la 4 o la División de Ingeniería en Ciencias de la Tierra, mientras que la división con la media más baja es 3 que corresponde a la División de Ingeniería Eléctrica.

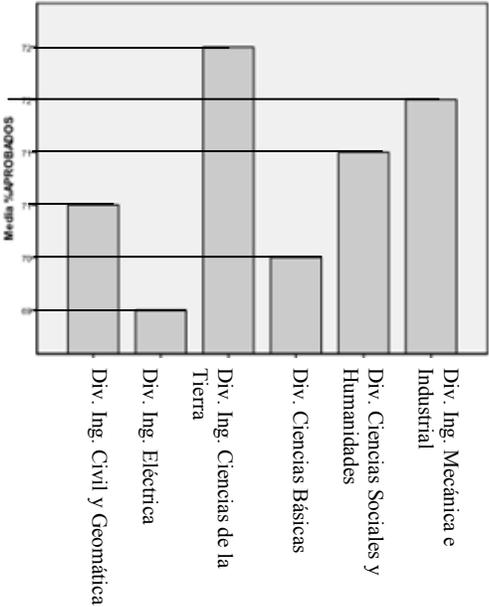


Figura 7.19 Gráfica de Desempeño por división, media de aprobados

En la figura 7.20 se muestra la media de la calificación del profesor por división, en la que se puede observar que el porcentaje de aprobados por grupo influye directamente en la calificación del profesor.

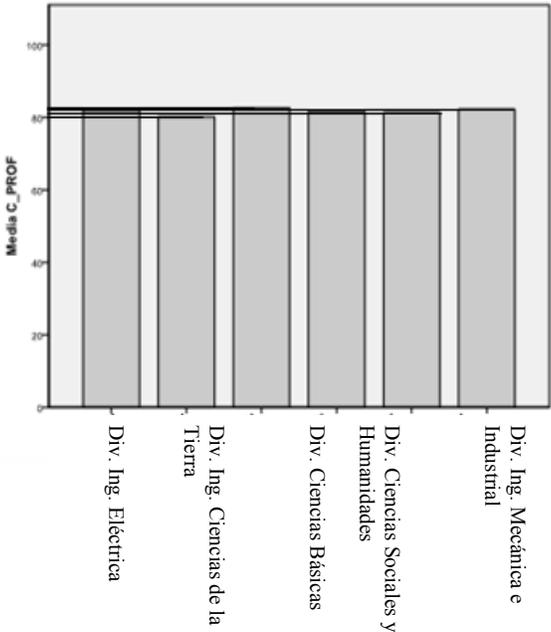


Figura 7.20 Gráfica de Desempeño por división, calificación del profesor

En la figura 7.21 se puede observar la media de la calificación del alumno por división, en la que se puede observar que la división con calificación del alumno más baja y más alta son la División de Ingeniería Mecánica e Industrial y la División de Ciencias Básicas, respectivamente.

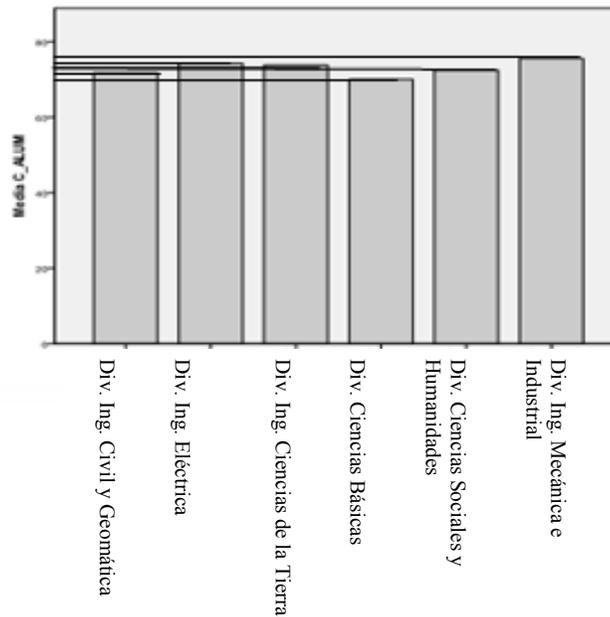


Figura 7.21 Gráfica de Desempeño por división, calificación del alumno

En la figura 7.22 se muestra la media de porcentaje de reprobados por división, en la que se puede observar que la división con mayor porcentaje de reprobados es la División de Ciencias Básicas.

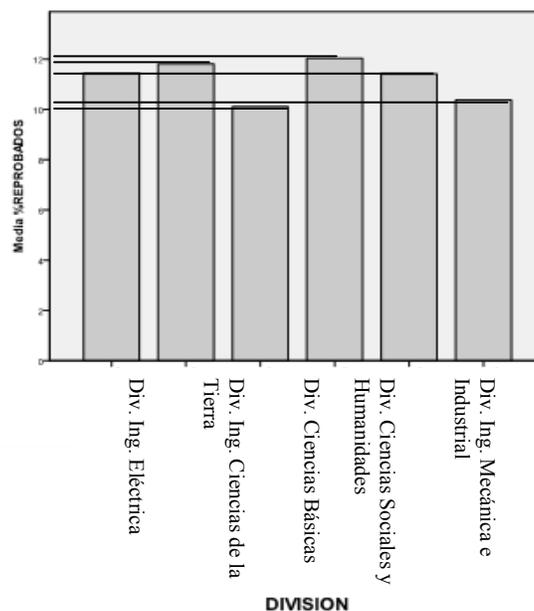
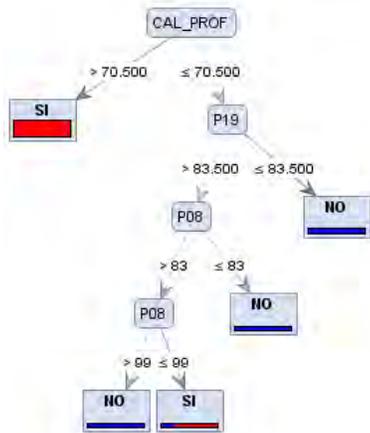


Figura 7.22 Gráfica de Desempeño por división, media de reprobados

7.2.2.4.1.1 División de Ingeniería Eléctrica

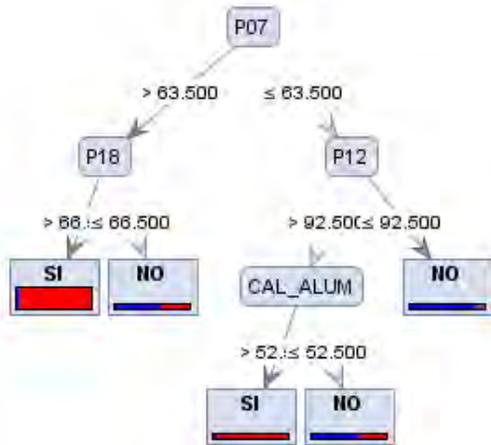
A continuación se realizó un ejemplo de estudio por división con árboles de decisión, los que se presentan son los árboles de decisión más representativos obtenidos mediante el algoritmo Random Forest para 20 árboles, para la división de Ingeniería Eléctrica. En el árbol de decisión 15 podemos observar que la mayoría de los profesores de la DIE tienen una calificación mayor a 70.5 y los que obtienen una calificación menor a 70.5 se debe a que en la pregunta 8 y en la pregunta 19 obtienen una calificación menor a 99 y a 83 respectivamente.



Nodo	Descripción
CAL_PROF	Calificación del profesor
P19	Volvería a tomar clases con el profesor
P08	El profesor promueve un ambiente de respeto en clase

Árbol de decisión 15. Calificación del Profesor

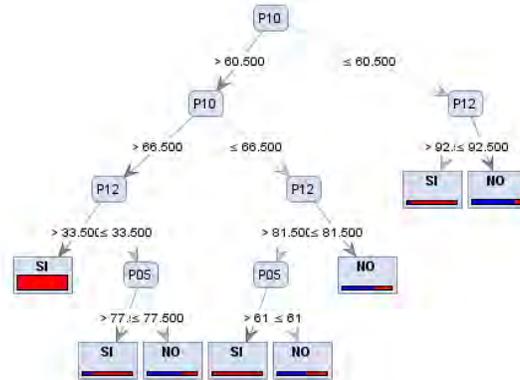
En el árbol de decisión 16, podemos observar la clasificación referente a la pregunta 7 en la que la mayoría de los profesores obtienen una calificación mayor a 63.5 y mayor a 66.5 en la pregunta 18.



Nodo	Descripción
P07	Propicia que los alumnos razonen sobre los temas que expone
P18	Se presentan ejemplos que facilitan la comprensión
P12	Regresa los exámenes, trabajos y proyectos corregidos

Árbol de decisión 16. Razonamiento de los Temas

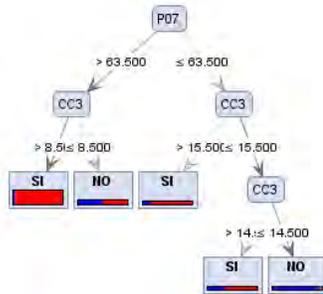
De la misma forma que en el árbol anterior, en el árbol de decisión 17 se observa la clasificación en relación a la pregunta 10 en la que los profesores que obtienen una calificación mayor a 80, obtienen a su vez una calificación mayor a 66.5 en la pregunta 10 y mayor a 33.5 en la pregunta 12.



Árbol de decisión 17. Resolución de Problemas

Nodo	Descripción
P10	Fomenta la habilidad de plantear y resolver problemas
P12	Regresa los exámenes, trabajos y proyectos corregidos
P05	Los conocimientos de profesor sobre la asignatura son

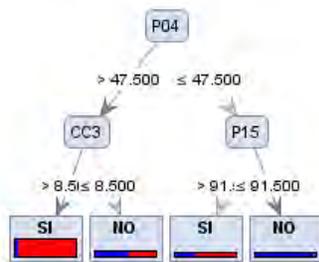
En el árbol de decisión 18 se observa que los profesores que obtienen una calificación mayor a 63.5 en la pregunta 7 obtienen una calificación mayor a 8.5 en la calificación de concepto CC3 y por lo tanto obtienen una calificación general mayor a 80.0.



Nodo	Descripción
P07	Propicia que los alumnos razonen sobre los temas que expone
CC3	Calificación de concepto de Evaluación

Árbol de decisión 18. Razonamiento de los Temas

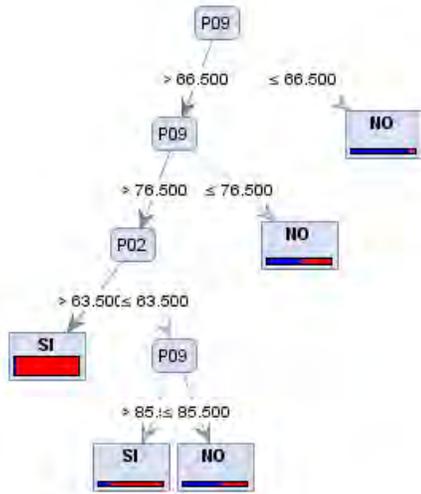
En el árbol de decisión 19 se observa que los profesores que obtienen una calificación mayor a 47.5 en la pregunta 4 y una calificación mayor a 80.5 en CC3, obtienen una calificación general mayor a 80 puntos.



Nodo	Descripción
P04	El profesor expone la clase con claridad
CC3	Calificación de concepto de Evaluación

Árbol de decisión 19. Exposición de la clase

En el árbol de decisión 20 se observa la clasificación en relación a la pregunta 9, y una de las ramas muestra que si su calificación en la pregunta 9 es mayor a 76.5 y en la pregunta 2 es mayor a 63.5 su calificación general será mayor a 80 puntos.



Nodo	Descripción
P09	Las tareas, trabajos y/o ejercicios que deja el profesor contribuyen al aprendizaje
P02	La confianza que el profesor inspira a los alumnos es

Árbol de decisión 20. Aprendizaje

En la figura 7.23 se observa la relación entre la calificación del alumno y la del profesor para la DIE.

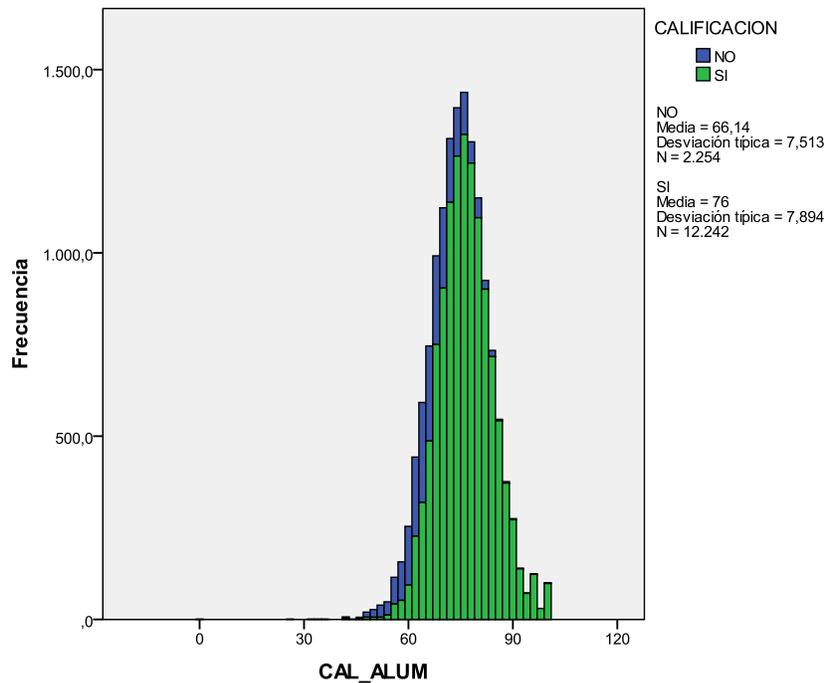
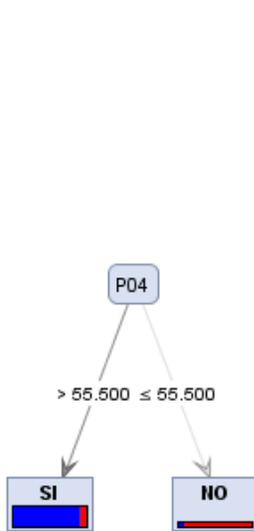


Figura 7.23 Gráfica de Desempeño de la DIE

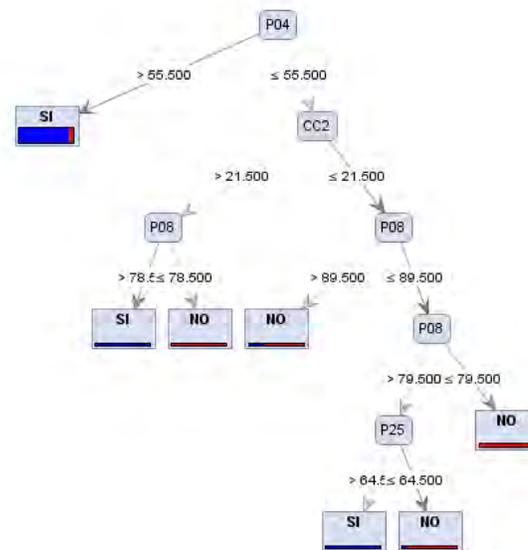
7.2.2.4.1.2 División de Ingeniería Civil

A continuación se presentan los árboles de decisión para la división de Ingeniería Civil.



Árbol de decisión 21 Exposición de la Clase

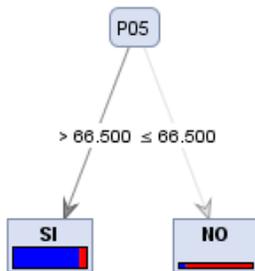
Nodo	Descripción
P04	Expone la clase en forma clara



Árbol de decisión 22. Exposición de la Clase

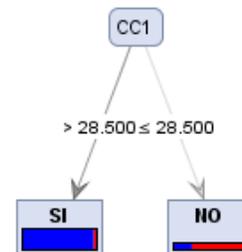
Nodo	Descripción
P04	Expone la clase en forma clara
CC2	Calificación de concepto de motivación
P08	Promueve un ambiente de respeto en clase
P25	Me siento motivado para estudiar la asignatura

En el árbol de decisión 21 se observa que la mayoría de los profesores de la División de Ingeniería Civil obtienen una calificación mayor a 55.5 en la pregunta 4, lo cual indica que exponen con claridad su clase, y a su vez que si un profesor expone con claridad su clase, obtendrá una calificación general mayor a 80; en el árbol de decisión 22 se observa que si un profesor de esta división no expone con claridad su clase, los alumnos no se sienten motivados para estudiar su asignatura.



Árbol de decisión 23. Conocimientos

Nodo	Descripción
P05	Conocimientos sobre la asignatura



Árbol de decisión 24. Aprendizaje

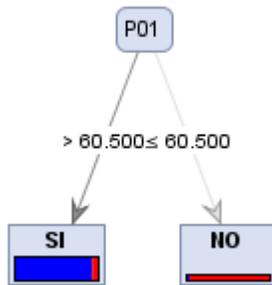
Nodo	Descripción
CC1	Calificación de concepto de aprendizaje

En el árbol de decisión 23 se observa que si un profesor tiene buenos conocimientos sobre su asignatura obtendrá una calificación general mayor a 80 en su evaluación y en el árbol

de decisión 24 se puede visualizar que si obtiene en su calificación de concepto de aprendizaje mayor a 28.5, obtendrá una calificación general mayor a 80.

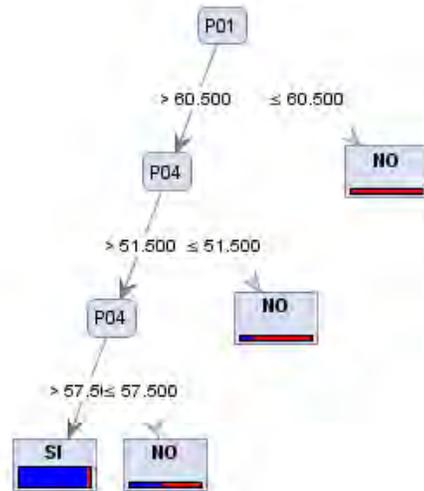
7.2.2.4.1.3 División de Ciencias Básicas

A continuación se presentan los árboles de decisión para la división de Ciencias Básicas.



Árbol de decisión 25 Interés

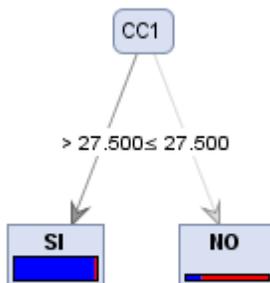
Nodo	Descripción
P01	Interés para que los alumnos aprendan



Árbol de decisión 26. Interés

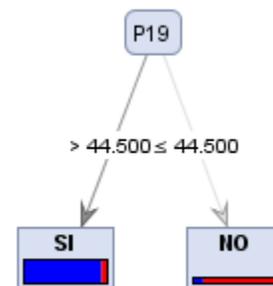
Nodo	Descripción
P01	Interés para que los alumnos aprendan
P04	Expone con claridad la clase

En los árboles de decisión 25 y 26 se puede observar la relación que existe entre la percepción que tiene el alumno sobre si su profesor tiene interés por que aprenda y si expone con claridad su clase, arrojando como resultado que si un profesor obtiene una calificación mayor a 60.5 en la pregunta 1 se debe a que también obtiene una buena calificación en la pregunta 4.



Árbol de decisión 27 Aprendizaje

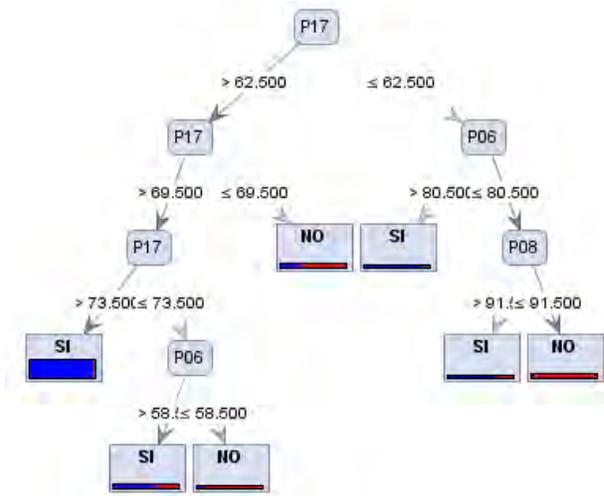
Nodo	Descripción
CC1	Calificación de concepto de aprendizaje



Árbol de decisión 28. Clases

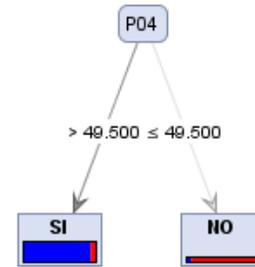
Nodo	Descripción
P19	Volvería a tomar clases con el profesor

En los árboles de decisión 27 y 28 se puede observar que si un profesor de esta división obtiene una calificación mayor a 27.5 y 44.5 en CC1 y en la pregunta 19 respectivamente, obtendrá una calificación general mayor a 80.



Árbol de decisión 29

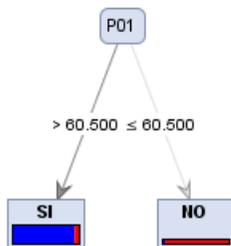
Nodo	Descripción
P17	La clase se da sin perder de vista los puntos a donde se quiere llegar
P06	Propicia que los alumnos amplíen por su cuenta sus conocimientos
P08	Promueve un ambiente de respeto en clase



Árbol de decisión 30

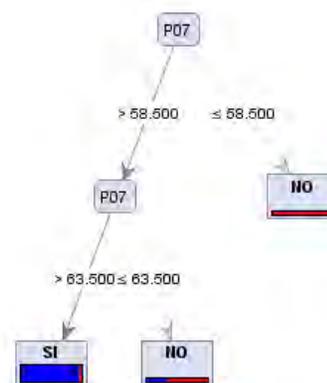
Nodo	Descripción
P04	Expone con claridad la clase

En el árbol de decisión 29 se puede observar que si un profesor da su clase sin perder de vista los puntos a donde quiere llegar también propicia que sus alumnos amplíen por su cuenta sus conocimientos, pero si un profesor obtiene en la pregunta 17 una calificación menor a 62.5, no promueve un ambiente de respeto en clase y no concreta los puntos a los que quiere llegar.



Árbol de decisión 31 Interés

Nodo	Descripción
P01	Interés para que los alumnos aprendan

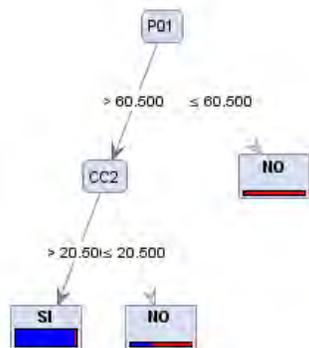


Árbol de decisión 32. Razonamiento

Nodo	Descripción
P07	Propicia que los alumnos razonen

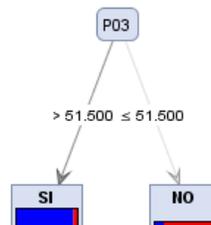
En el árbol de decisión 31 se observa que la mayoría de los profesores de esta división obtienen una calificación mayor a 60.5 en la pregunta 1 y a su vez mayor a 80 en su

calificación general; en el árbol de decisión 32 se observa el mismo análisis pero realizado respecto a la pregunta 17.



Árbol de decisión 33 Motivación

Nodo	Descripción
P01	Interés para que los alumnos aprendan
CC2	Calificación de concepto de motivación



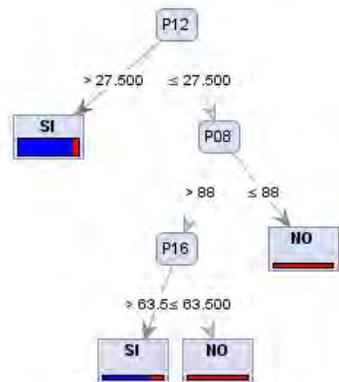
Árbol de decisión 34. Clases interesantes

Nodo	Descripción
P03	La clase se desarrolla de manera interesante

En el árbol de decisión 33 se observa que si un profesor tiene interés por que los alumnos aprendan los alumnos se sentirán motivados para estudiar la asignatura; en el árbol de decisión 34 se observa que la mayoría de los profesores de esta división obtienen una calificación mayor a 51.5 en la pregunta 3.

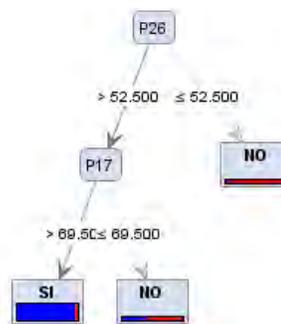
7.2.2.4.1.4 División de Ciencias Sociales y Humanidades

A continuación se presentan los árboles de decisión para la división de Ciencias Sociales y Humanidades.



Árbol de decisión 35. Evaluación

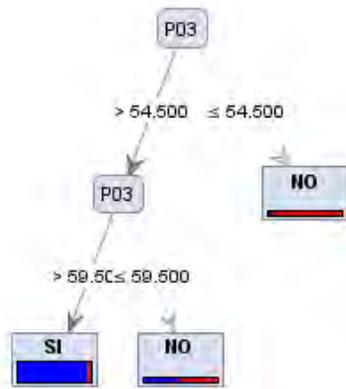
Nodo	Descripción
P12	Regresa los exámenes, tareas, trabajos y proyectos corregidos
P08	Promueve un ambiente de respeto en clase
P16	Asistencia a clase



Árbol de decisión 36. Aprendizaje

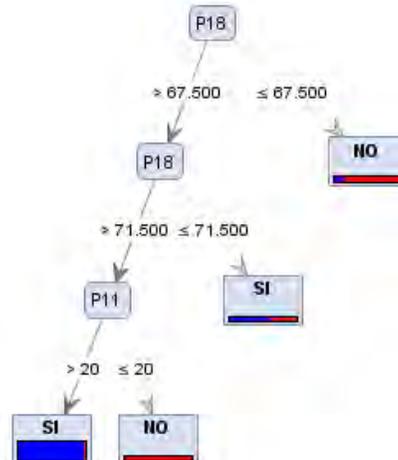
Nodo	Descripción
P26	Aprendizaje de los temas
P17	La clase se da sin perder de vista los puntos a los que se quieren llegar

En árboles de decisión 35 y 36 se observa la relación que existe entre la pregunta 12, 16 y 8, y entre la pregunta 26 y 17 respectivamente. El árbol 35 se refiere a la constancia del profesor en su clase y el árbol 36 se refiere al aprendizaje de los alumnos.



Árbol de decisión 37. Interés

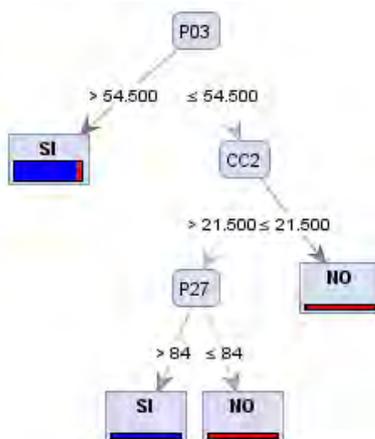
Nodo	Descripción
P03	La clase se desarrolla de manera interesante



Árbol de decisión 38. Comprensión

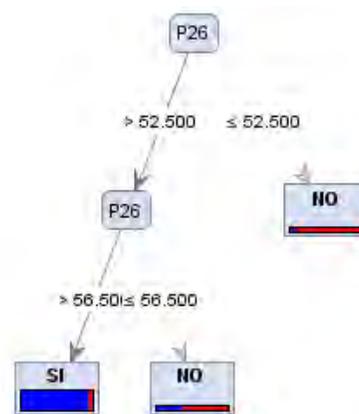
Nodo	Descripción
P18	Se presentan ejemplos que facilitan la comprensión
P11	Puntualidad al llegar al salón de clase

En los árboles de decisión 37 y 38 se muestra el comportamiento de los profesores de la división de Ciencias Sociales y Humanidades respecto a la pregunta 3 y la pregunta 18 respectivamente.



Árbol de decisión 39. Motivación

Nodo	Descripción
P03	Las clases se desarrollan de manera interesante
CC2	Calificación de concepto de motivación
P27	Realizo tareas y trabajos asignados



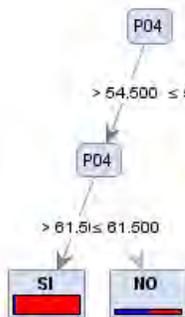
Árbol de decisión 40. Aprendizaje

Nodo	Descripción
P26	Mi aprendizaje de los temas ha sido

En árboles de decisión 39 y 40 de la misma forma que en los árboles anteriores, se puede observar el desarrollo de los árboles de decisión respecto a las preguntas 3 y 26 respectivamente.

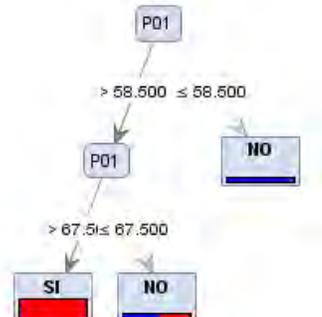
7.2.2.4.1.5 División de Ingeniería Mecánica e Industrial

A continuación se presentan los árboles de decisión para la división de Ingeniería Mecánica e Industrial.



Árbol de decisión 41. Exposición

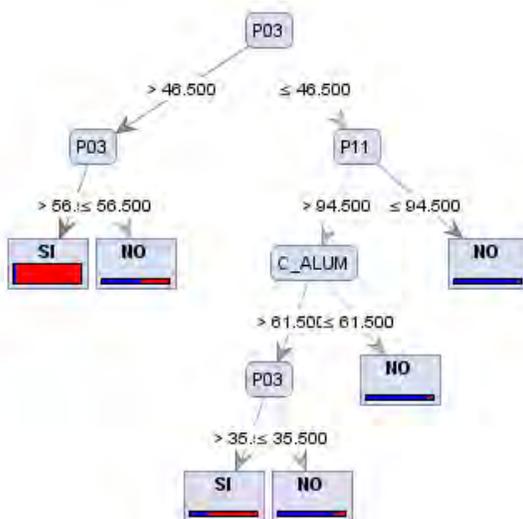
Nodo	Descripción
P04	Expone con claridad la clase



Árbol de decisión 42. Interés

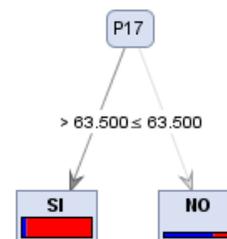
Nodo	Descripción
P01	Interés para que los alumnos aprendan

En los árboles de decisión 41 y 42 se muestra que los profesores de la división de ingeniería mecánica e industrial obtienen una calificación menor a 54.5 en la pregunta 4 y menor a 59.5 en la pregunta 1, respectivamente.



Árbol de decisión 43. Puntualidad

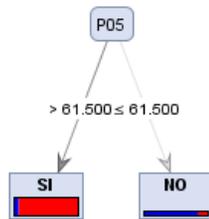
Nodo	Descripción
P03	La clase se desarrolla de manera interesante
P11	Puntualidad al llegar al salón de clase
C_ALUM	Calificación del alumno



Árbol de decisión 44. Desarrollo de la Clase

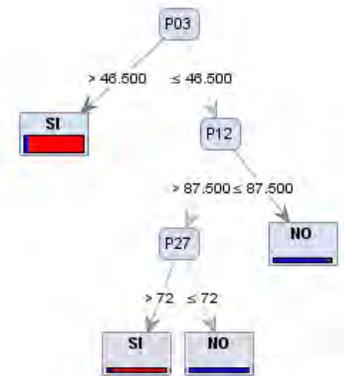
Nodo	Descripción
P17	La clase se da sin perder de vista los puntos a los que se quiere llegar
P27	Realizo tareas y trabajos asignados

En el árbol de decisión 43 se observa el comportamiento de los profesores de esta división en relación a la pregunta 3, mientras que en el árbol 44 se observa que la mayoría de los profesores tienen una calificación mayor a 63.5 en la pregunta 17.



Árbol de decisión 45. Conocimientos

Nodo	Descripción
P05	Conocimientos sobre la asignatura
P08	Promueve un ambiente de respeto en clase
P16	Asistencia a clase



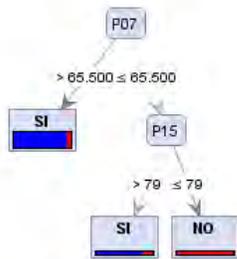
Árbol de decisión 46. Desarrollo de la Clase

Nodo	Descripción
P03	La clase se desarrolla de manera interesante
P12	Regresa los exámenes, tareas, trabajos y proyectos corregidos

En el árbol de decisión 45 se puede observar que la mayoría de los profesores de esta división obtienen una calificación mayor a 61.5 en la pregunta 5 que se refiere a los conocimientos del profesor; en el árbol de decisión 46 se observa que un profesor que no desarrolla sus clases de forma interesante, al mismo tiempo, no regresa los trabajos corregidos.

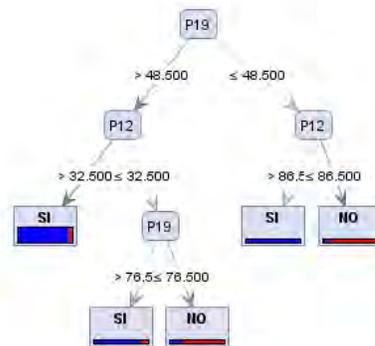
7.2.2.4.1.6 División de Ciencias de la Tierra

A continuación se presenta el mismo análisis realizado anteriormente pero para las demás divisiones de la Facultad de Ingeniería. Se continúa con la División de Ciencias de la Tierra.



Árbol de decisión 47. Razonamiento

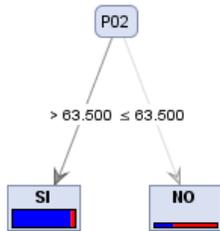
Nodo	Descripción
P07	El profesor propicia que los alumnos razonen
P15	Entrega oportunamente las calificaciones



Árbol de decisión 48. Evaluación

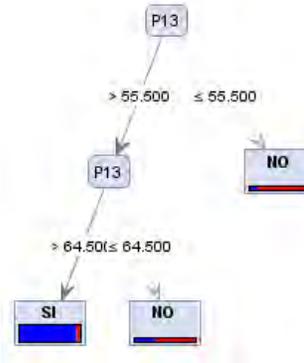
Nodo	Descripción
P19	Volvería a tomar clases con el profesor
P12	Regresa los exámenes, proyectos y trabajos corregidos

En los árboles de decisión 47 y 48, se puede observar que la mayoría de los profesores de la división de Ciencias de la Tierra, obtienen una calificación mayor a 67 y mayor a 48.5 en las preguntas 7 y 19 respectivamente; además se puede observar que los resultados de estas preguntas, están relacionados con las preguntas 15 y 12 y 19 respectivamente.



Árbol de decisión 49. Confianza

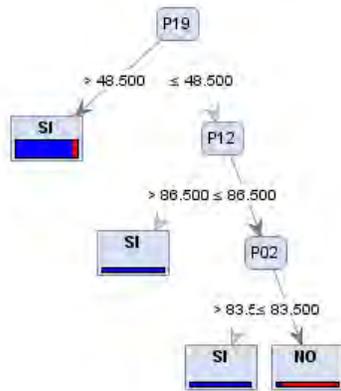
Nodo	Descripción
P02	La confianza que el profesor inspira a los alumnos



Árbol de decisión 50. Evaluación

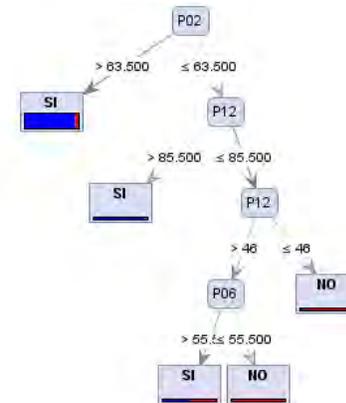
Nodo	Descripción
P13	Los exámenes y/o evaluaciones tienen un grado de dificultad adecuado

En los árboles de decisión 49 y 50 se puede observar las calificaciones de los profesores respecto a la pregunta 2 y 13; en la pregunta 2 se visualiza que la mayoría de los profesores correspondientes a esta división tienen una calificación mayor a 63.5 y en la pregunta 13 se observa que la mayoría obtiene una calificación mayor a 64.5 en la pregunta 13.



Árbol de decisión 51. Confianza

Nodo	Descripción
P19	Volvería a tomar clases con el profesor
P02	La confianza que el profesor inspira es
P12	Regresa los exámenes, proyectos y trabajos corregidos



Árbol de decisión 52. Confianza

Nodo	Descripción
P02	La confianza que el profesor inspira es
P12	Regresa los exámenes, proyectos y trabajos corregidos
P06	El profesor propicia que los alumnos amplíen por su cuenta sus conocimientos

En el árbol de decisión 51 se observa que si un alumno de la división de Ciencias de la Tierra, elige que no volvería a tomar clases con el profesor se debe a que el profesor inspira poca confianza a sus alumnos a pesar de que cumple con la entrega de trabajos y proyectos corregidos; en el árbol de decisión 52 se observa que si un profesor inspira poca confianza a sus alumnos se debe a que no entrega los trabajos corregidos y no propicia que sus alumnos amplíen por su cuenta sus conocimientos sobre la asignatura.

Dentro de la encuesta debe ser considerada, la cantidad de alumnos inscritos y los alumnos encuestados, ya que debido a lo anterior podemos decir si la encuesta es objetiva o no. A continuación se presenta una gráfica en que se puede observar la cantidad de alumnos inscritos en relación a los alumnos encuestados.

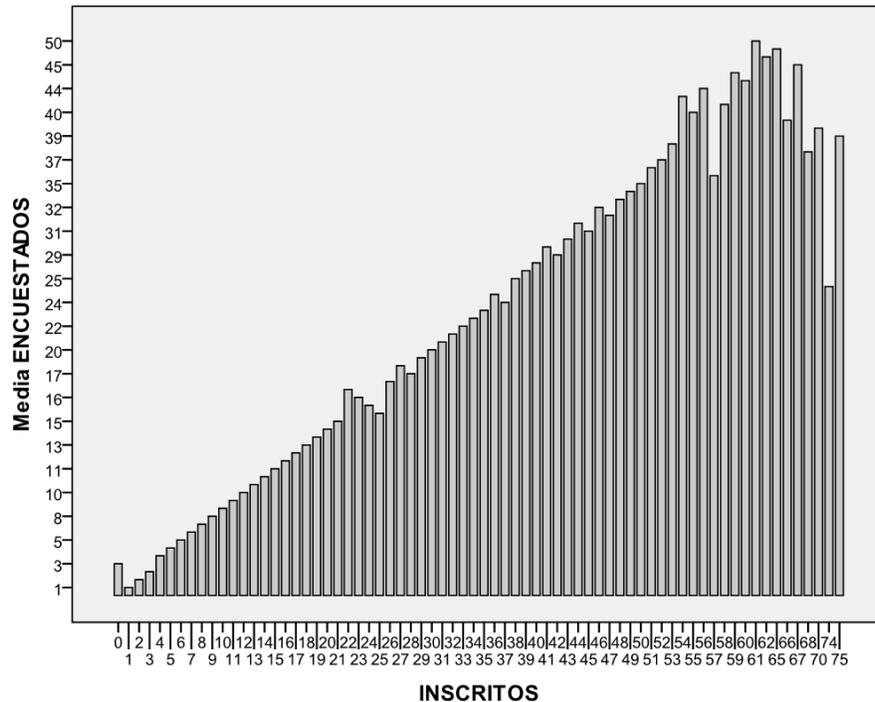


Figura 7.24 Gráfica de encuestados e inscritos de la DIE

7.2.2.4.2 Análisis por Departamento

A continuación se muestra un ejemplo, haciendo el análisis por departamento de la DIE, para que se pueda observar cómo se puede hacer el estudio de minería de datos, partiendo de lo más general, hacia lo particular.

La división de Ingeniería Eléctrica, cuenta con los siguientes departamentos y sus respectivas claves:

- Departamento de Ingeniería en Computación 334
- Departamento de Ingeniería de Control 332
- Departamento de Ingeniería Eléctrica de Potencia 331
- Departamento de Ingeniería Electrónica 333
- Departamento de Sistemas Energéticos 336
- Departamento de Ingeniería en Telecomunicaciones 335

En las figuras 25 y 26 se muestra el desempeño por departamento de la DIE, considerando la calificación del alumno en función de la calificación de su profesor.

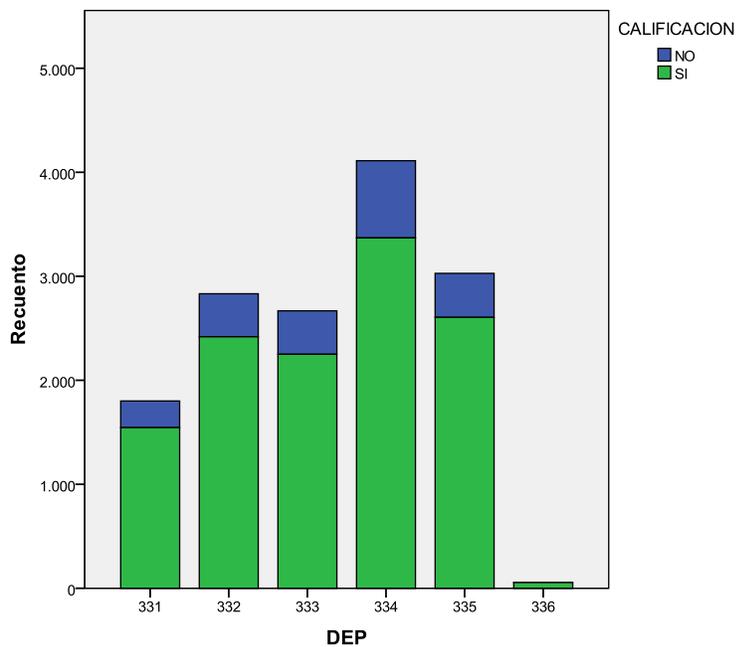


Figura 7.25 Gráfica de Desempeño de la DIE por departamento

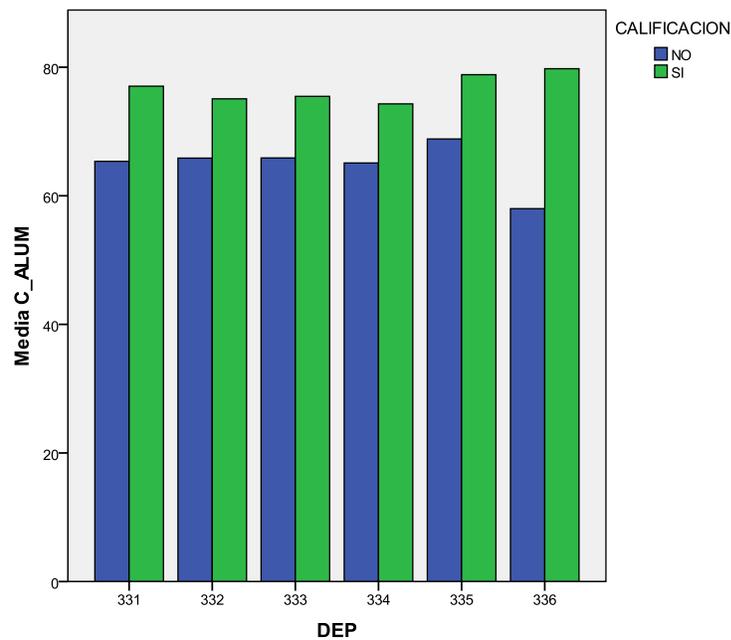


Figura 7.26 Gráfica de Calificaciones de Alumnos de la DIE

A continuación se muestra un agrupamiento por medio del algoritmo K-Means para observar las características del departamento 333 que se refiere al departamento de Ingeniería Electrónica, mediante el software SPSS v.18., los resultados se observan en la tabla 7.12.

Centros de los conglomerados finales					
	Conglomerado				
	1	2	3	4	5
DEP	333	333	333	333	333
C_PROF	67	79	51	91	79
P01	68	84	51	91	77
P15	49	49	32	88	79
P14	80	88	68	93	89
P13	69	81	52	91	80
P10	64	82	46	90	75
P12	46	45	29	87	78
P11	74	85	56	94	87
P03	60	77	45	86	69
P09	75	88	60	94	84
P08	82	92	71	95	88
P07	68	84	52	91	77
P02	69	84	52	91	77
P06	66	79	53	87	74
P05	73	87	59	92	80
P04	62	83	42	91	73

Tabla 7.12 Resultados de Agrupamiento

En la tabla 7.13 se muestra un cuadro comparativo de los árboles de decisión realizados por división que son de la misma pregunta o concepto, lo anterior con el fin de poder comparar el desempeño de cada división respecto a la pregunta o concepto a la que se refieren los árboles; cabe señalar que los árboles de las divisiones son diferentes entre sí y sólo algunos árboles coinciden en el nodo raíz, es por eso que no se mencionan todas las divisiones y se vuelve necesario indicar a que división pertenece cada árbol.

Se puede observar que en la primera fila de la tabla 7.13 se menciona el nombre de la división a la que pertenece el árbol y se observan los árboles que tienen como nodo raíz la pregunta 7 que corresponde a si el profesor propicia que los alumnos razonen; se puede observar que para la División de Ciencias Básicas su promedio en esta pregunta es de 63.5, mientras que para la División de Ciencias de la Tierra es 65.5 y para la División de Ingeniería Eléctrica es 63.5. Se puede realizar la misma comparación para las siguientes filas.

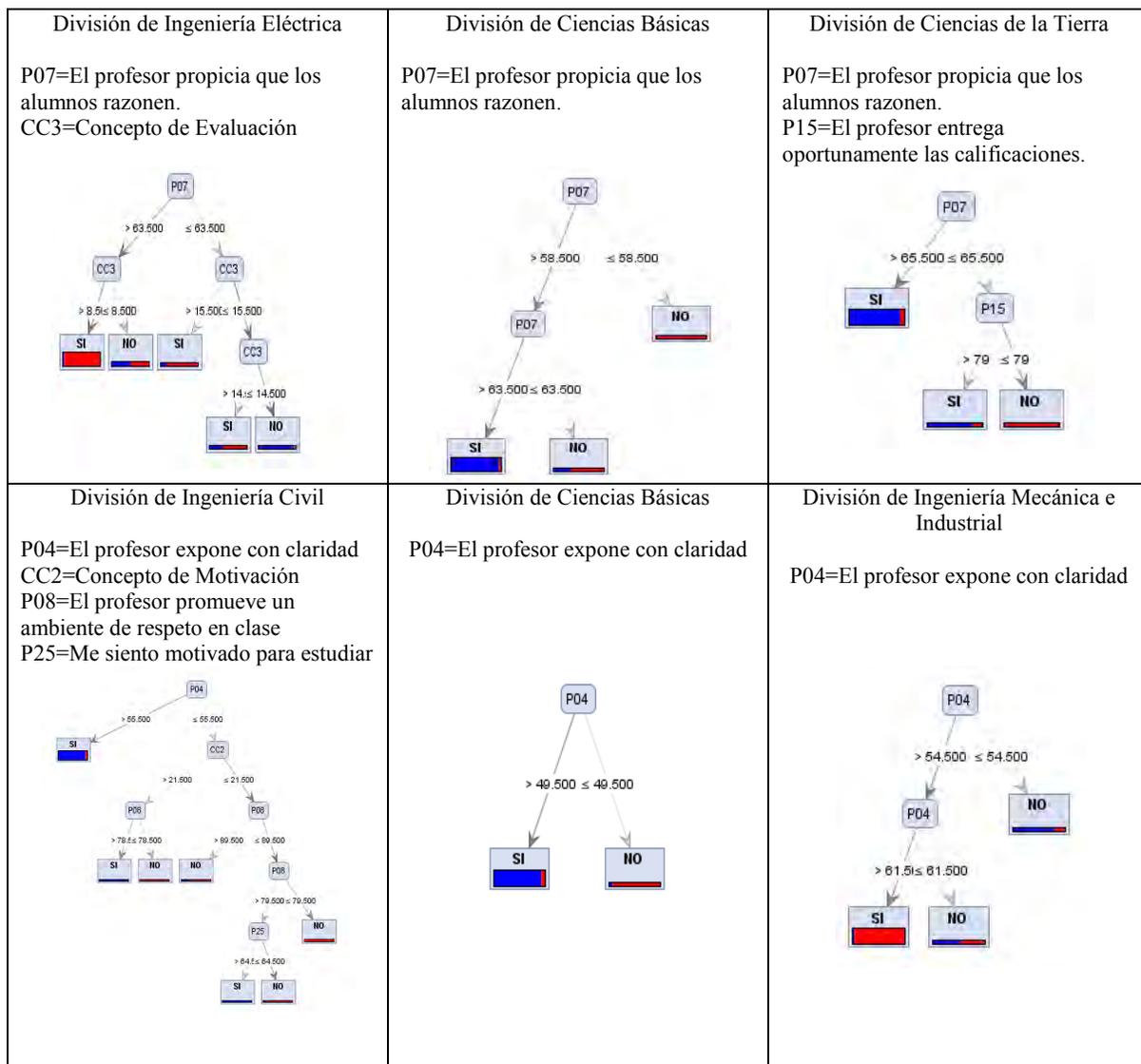
Para una mejor toma de decisión se determina que los profesores de la División de Ciencias de la Tierra y la División de Ingeniería Eléctrica necesitan tomar cursos de evaluación del aprendizaje escolar; si por ejemplo se tratara de los profesores que corresponden a los árboles de la pregunta 4 (Expone con claridad la clase), se determinaría que dichos profesores requieren mejorar su expresión verbal y por lo tanto remitirlos a los cursos correspondientes.

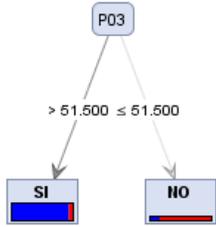
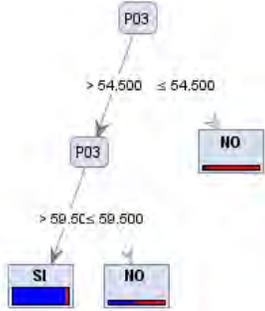
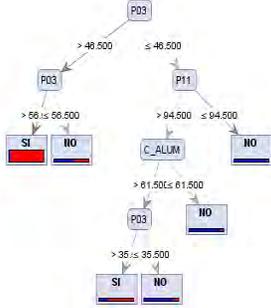
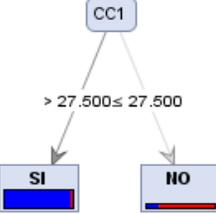
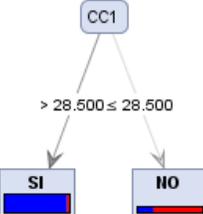
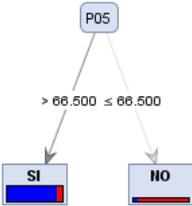
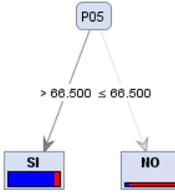
Otra toma de decisión se puede observar en los árboles contenidos en la tabla 7.13 referentes a la pregunta 3 que habla sobre si las clases se desarrollan de manera interesante. En esta pregunta destacan las divisiones de Ciencias Básicas, Sociales y Humanidades y Mecánica e Industrial; sobresale que las tres divisiones tienen una calificación que va desde el 46.500 hasta el 54.500, por lo tanto en una toma de decisión se sugeriría que los

profesores pertenecientes a estas divisiones asistan a cursos sobre didáctica o expresión oral.

También se puede observar que en la última fila de la tabla 7.13 los árboles hacen referencia a la pregunta 1 en la que se menciona el interés del profesor por que los alumnos aprendan. Las divisiones que generaron árbol respecto a esta pregunta fueron: división de Ciencias Básicas y la división de Ingeniería Mecánica e Industrial; estas divisiones tienen un promedio en estas preguntas de 60.500 y de 58.500 respectivamente, se podría sugerir que a estos profesores se les oriente sobre cursos de expresión oral y corporal.

Respecto a los árboles en los que se hace referencia a la pregunta 5 que pregunta si los conocimientos del profesor son buenos, se muestra la división de Ingeniería Mecánica e Industrial y la división de Ingeniería Civil, se observa que ambas divisiones tienen 66.500 en promedio para esta pregunta, por lo que podría sugerirse cursos de actualización sobre la asignatura que imparten los profesores de estas divisiones.



<p>División de Ciencias Básicas</p> <p>P03=Las clases se desarrollan de manera interesante.</p> 	<p>División de Ciencias Sociales y Humanidades</p> <p>P03=Las clases se desarrollan de manera interesante.</p> 	<p>División de Ingeniería Mecánica e Industrial</p> <p>P03=Las clases se desarrollan de manera interesante. P11=El profesor llega al salón de clase C_ALUM=Calificación general del alumno</p> 
<p>División de Ingeniería Civil</p> <p>CC1=Concepto de Aprendizaje</p> 	<p>División de Ciencias Básicas</p> <p>CC1=Concepto de Aprendizaje</p> 	
<p>División de Ingeniería Mecánica e Industria</p> <p>P05=Los conocimientos del profesor sobre la asignatura son</p> 	<p>División de Ingeniería Civil</p> <p>P05=Los conocimientos del profesor sobre la asignatura son</p> 	

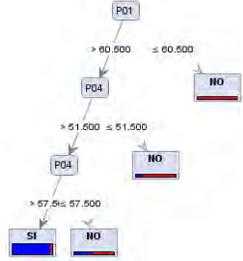
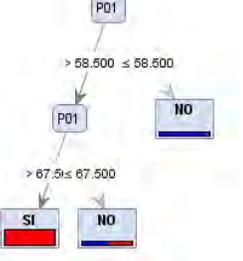
<p>División de Ciencias Básicas</p> <p>P01=El interés del profesor por que los alumnos aprendan es P04=El profesor expone con claridad la clase</p> 	<p>División de Ingeniería Mecánica e Industrial</p> <p>P01=El interés del profesor por que los alumnos aprendan es</p> 	
---	--	--

Tabla 7.13 Comparación de Divisiones

7.2.2.5 Agrupamiento o Clustering

Otra forma de obtener características relevantes de los datos es a través del Agrupamiento o Clustering, la cual consiste en agrupar un cúmulo de datos dependiendo de sus características principales. Esta tarea de Minería de Datos se puede realizar a través de Rapid Miner; al igual que el procedimiento anterior, se debe de seleccionar una fuente de datos, una tarea y realizar las conexiones necesarias en el diagrama de bloques. En la figura 7.27 se muestra el diagrama de bloques del proceso de agrupamiento, trabajando con el algoritmo K-Means.

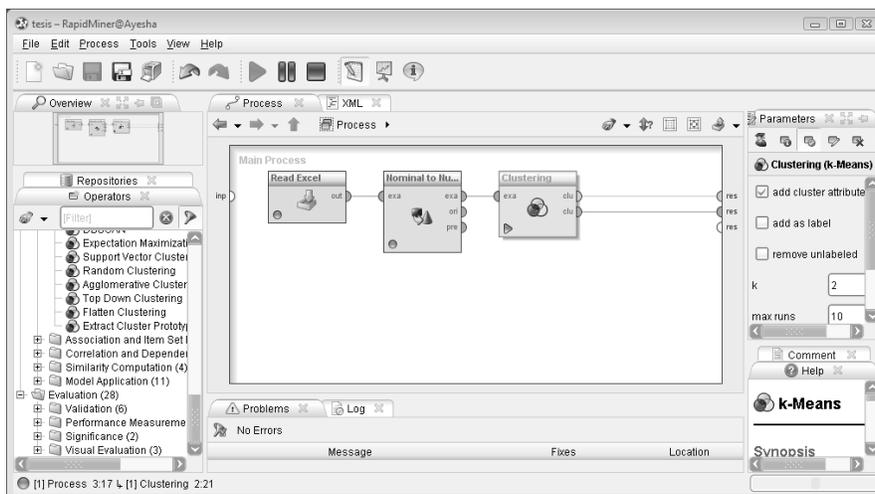


Figura 7.27 Diagrama de Bloques de Agrupamiento en Rapid Miner

Al ejecutar el diagrama de bloques anterior, se obtienen 10 clústers en los que cada uno contiene características diferentes. En la figura 7.28 se muestra la pantalla de resultados de Rapid Miner.

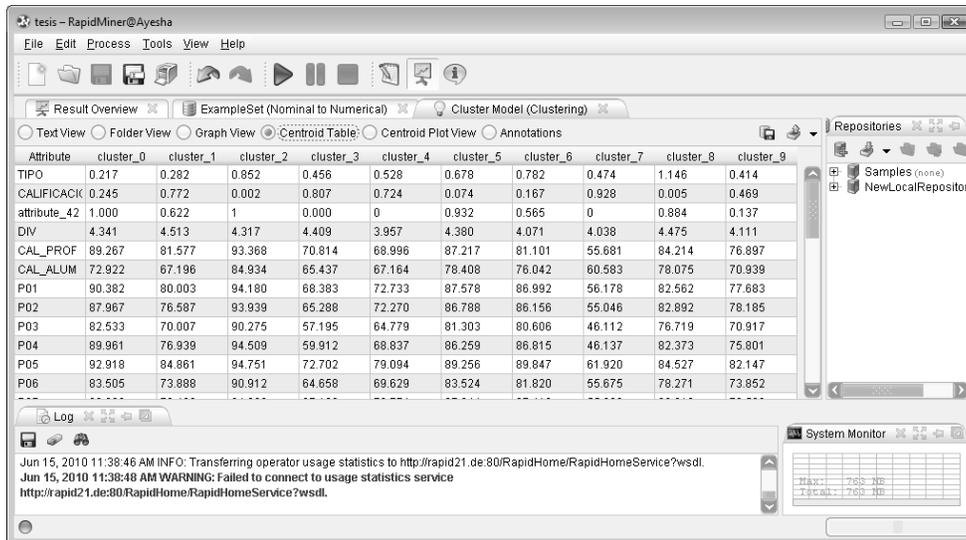


Figura 7.28 Resultados de Agrupamiento en Rapid Miner

En este tipo de tarea, tenemos la opción de ver de diferentes formas el resultado; el primero es el modo Text View en el que se puede observar los clústers, realizados y la cantidad de datos que cada uno contiene, tal y como se muestra en la figura 7.29.

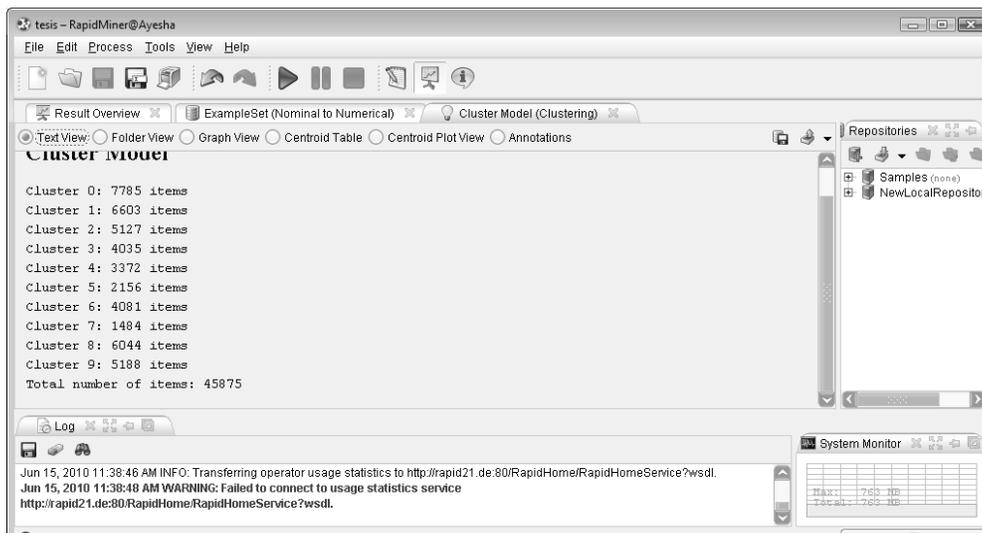


Figura 7.29 Resultados de Agrupamiento en modo Text View

Otra forma de visualizar los resultados de Agrupamiento en Rapid Miner, es a través de la opción Folder View, en el que se pueden observar las pruebas para cada clúster y sus valores correspondientes para cada uno de los atributos como se muestra en la figura 7.30.

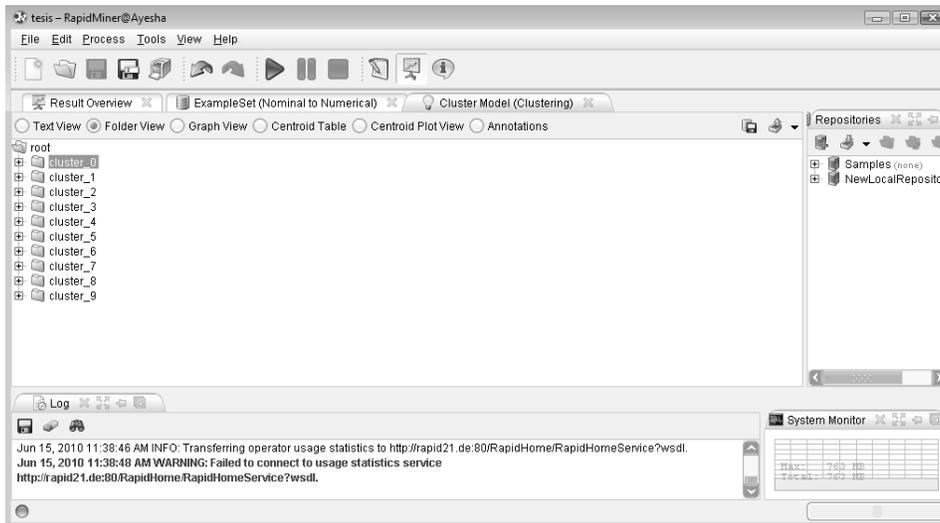


Figura 7.30 Resultados de Agrupamiento en modo Folder View

Uno de los modos que ayuda a visualizar la composición de los agrupamientos es Centroid Plot View, en el que se puede observar gráficamente como se distribuyen los valores para cada grupo, con el algoritmo K-Means, como se muestra en la figura 7.31.

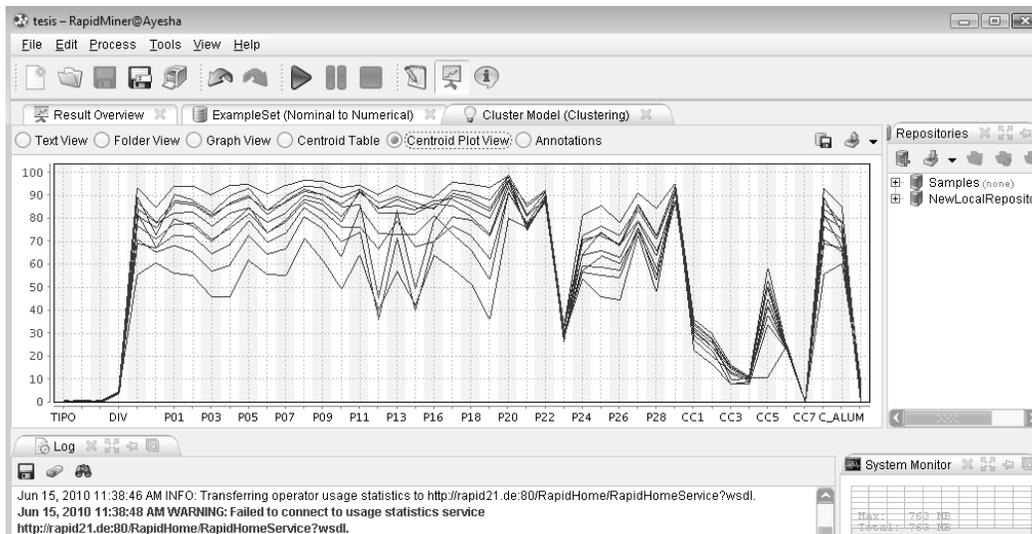


Figura 7.31 Resultados de Agrupamiento en modo Centroid Plot View

7.2.2.6 Redes Neuronales

A continuación se describe la creación de una red neuronal mediante RapidMiner, utilizando el algoritmo de perceptrón multicapa, basada principalmente en la clasificación de profesores en base a su calificación general de las encuestas y su porcentaje de alumnos aprobados y reprobados. Fue utilizado el algoritmo perceptrón multicapa, el cual es uno de los algoritmos más conocidos, precisos y económicos en cuanto a tiempo de ejecución para

redes neuronales; a continuación se muestran los resultados obtenidos con este método. Cabe destacar que la red neuronal tiene como datos de entrada para la capa de entrada las preguntas válidas descritas en este capítulo (pp. 82) y los alumnos con 5, NP y aprobados por grupo.

En las figuras 7.32 y 7.33 se puede observar la red neuronal obtenida.

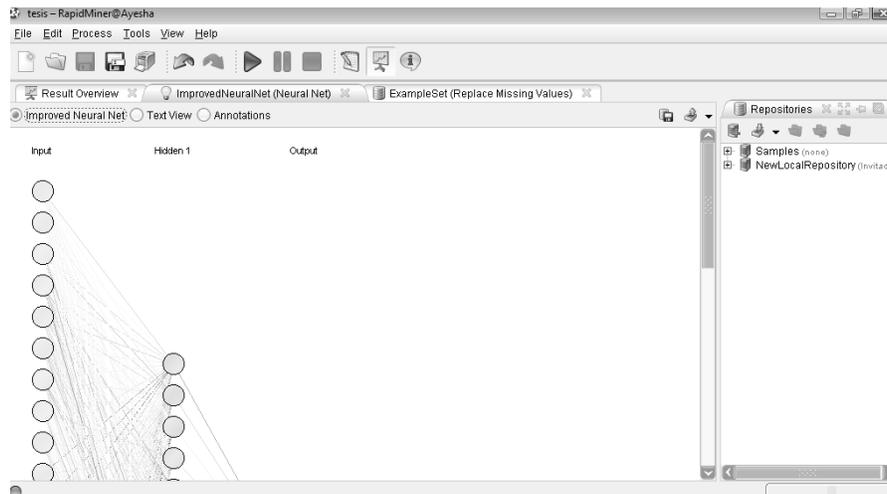


Figura 7.32 Red Neuronal a.

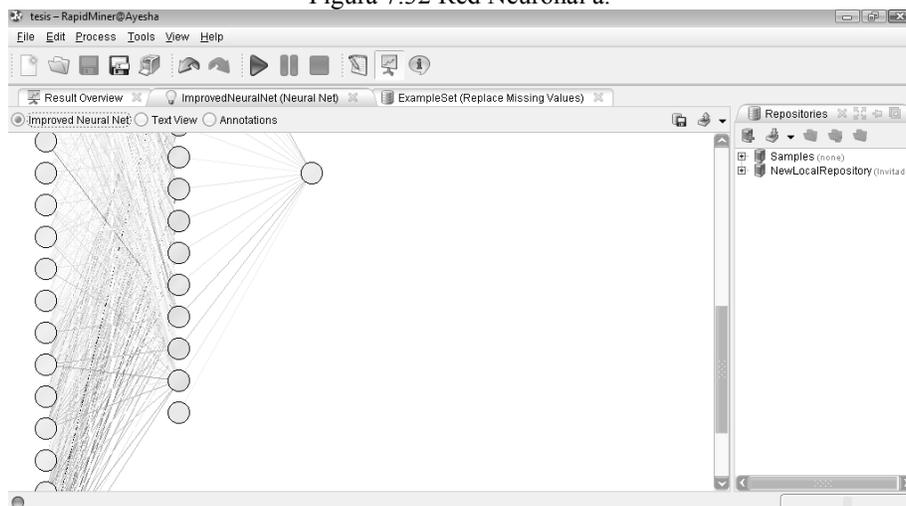


Figura 7.33 Red Neuronal b.

En las figuras 7.32 y 7.33, se puede observar que esta red neuronal cuenta con una capa oculta que se encuentra entre la capa de entrada y la capa de salida; las operaciones que se realizan en cada nodo, fueron descritas en el capítulo 3 de esta tesis correspondiente a Minería de Datos.

En la figura 7.34, se puede observar los resultados del proceso anterior.

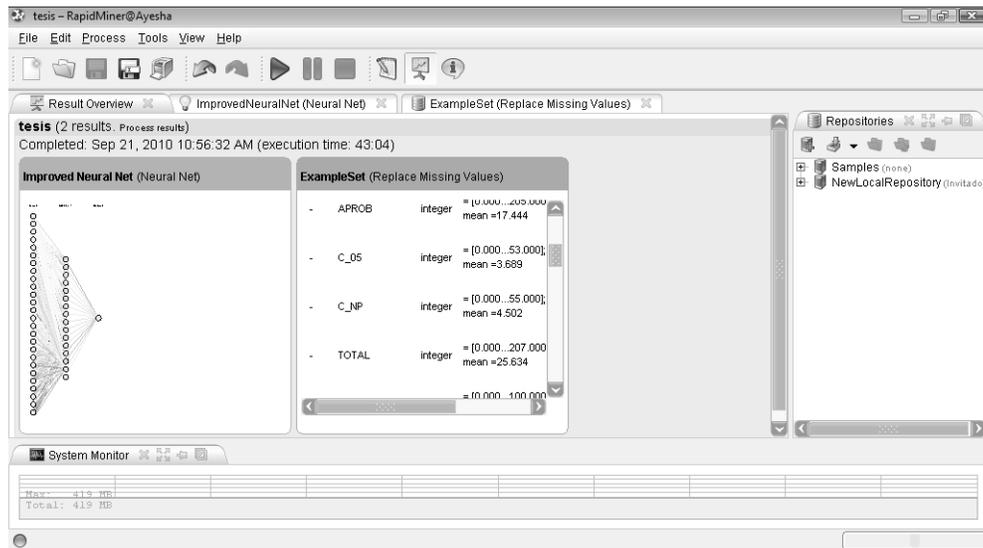


Figura 7.34 Resultados de RapidMiner

En la figura 7.35 se pueden observar los metadatos del proceso anterior.

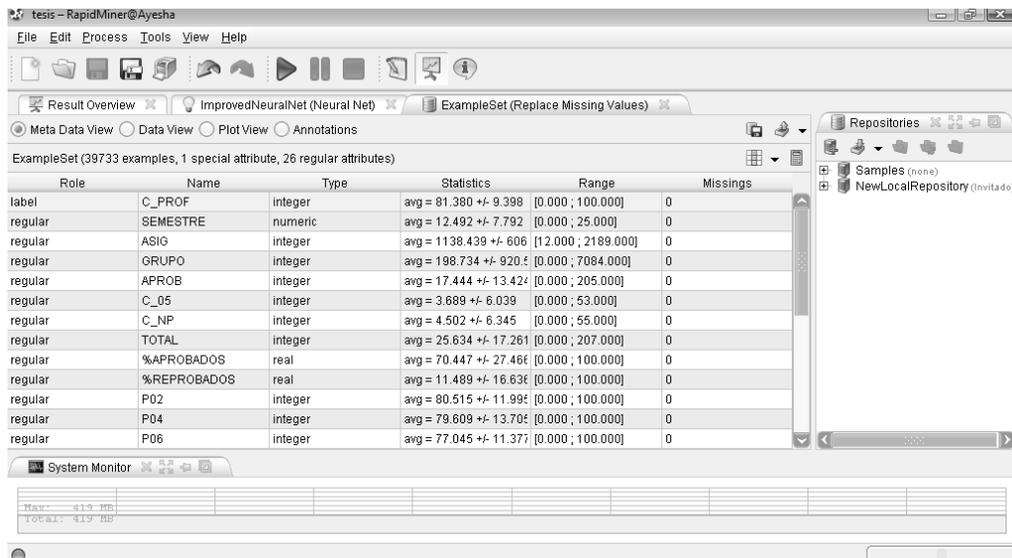


Figura 7.35 Metadatos RapidMiner

La clasificación que realiza la red neuronal es muy parecida a los árboles de decisión o a los clúster pero es más precisa mediante el presente método ya que se realiza a través uno de los algoritmos más eficientes de la inteligencia artificial.

En la figura 7.36 se muestra el proceso de clasificación que realiza la red neuronal obtenida con anterioridad.

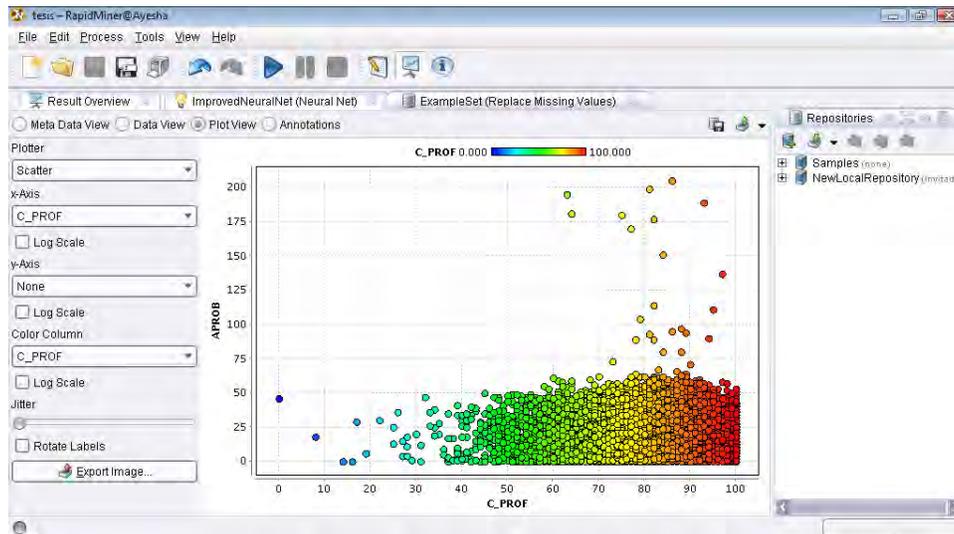


Figura 7.36 Clasificación de calificación del profesor

En la figura 7.37 se muestran los resultados de clasificación de la calificación de los alumnos.

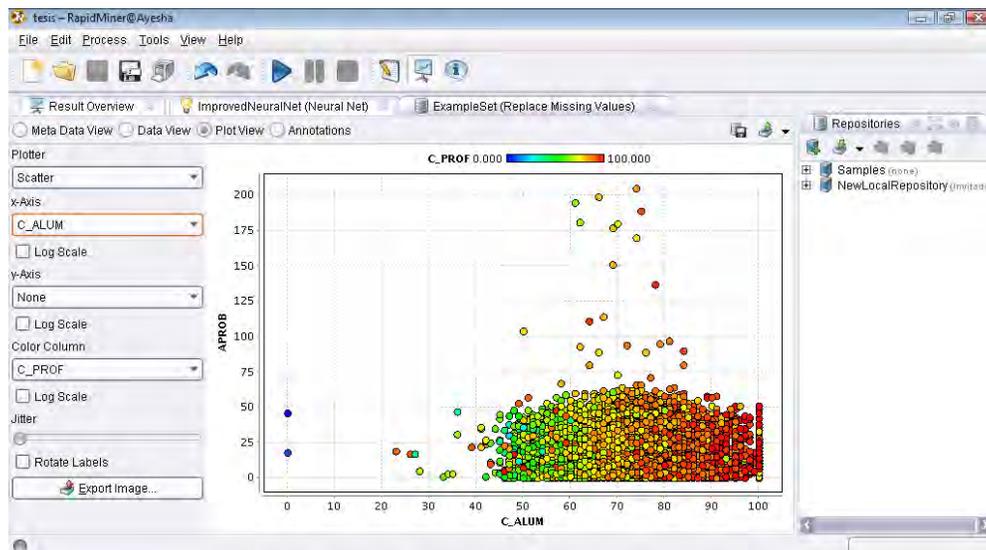


Figura 7.37 Clasificación de calificación del alumno

RapidMiner nos permite observar la clasificación de la Red Neuronal para cada una de las variables introducidas.

El proceso de Minería de Datos también se puede llevar a cabo a través de WEKA el cual nos permite realiza la consulta desde su ventana de OpenDB, la cual tiene la tarea de conectar la base de datos al software de minería de datos, en este caso OpenDB, lo cual nos permite realizar Minería Relacional; el término anterior, se refiere a un proceso de Minería

de Datos que obtiene la información necesaria para su proceso, mediante una consulta a la Base de Datos que incluye dos o más tablas.

A través de WEKA y OpenDB se realizó la siguiente consulta:

```
SELECT e.*, p.nom_prof, p.rfc, g.grupo
where encuesta e, profesor p, grupo g
and p.rfc=g.rfc
and g.grupo=e.grupo
and num_rows<1000
```

La consulta anterior se realizó con el fin de obtener los datos correspondientes a la encuesta, es decir, 29 preguntas, división, grupo, departamento, calificación del alumno, calificación del profesor, etc; y de esta forma poder realizar un proceso de minería de datos semejante al anteriormente descrito pero solicitando información directamente a la base de datos y de forma relacional.

7.3 Manual del Usuario

7.3.1 Descargar RapidMiner

Para poder hacer uso de la herramienta de minería de datos RapidMiner es necesario realizar la descarga del software de la página oficial de RapidMiner que se encuentra en la dirección siguiente: <http://rapid-i.com>. Esta página únicamente solicita al usuario, ingresar algunos datos personales para poder realizar la descarga, ya que como se menciona en capítulos anteriores, RapidMiner es software libre; debido a que RapidMiner cuenta con diferentes versiones, el usuario es capaz de seleccionar el software que más le sea conveniente dependiendo de su sistema operativo. En la figura 7.38 se puede observar la página principal de RapidMiner.



Figura 7.38 Página de Bienvenida de RapidMiner

Una vez descargado el software, se procede a la instalación del mismo, que consiste en seguir una serie de pasos que llevan a una instalación exitosa. Una vez instalado el programa, aparece un ícono en el escritorio como se muestra en la figura 7.39.



Figura 7.39 Pantalla de ícono de RapidMiner

7.3.2 Iniciar RapidMiner

Para iniciar RapidMiner basta con dar doble click sobre el ícono mostrado en la figura anterior o ingresar al menú de inicio de Windows, dar click en *todos los programas*, y seleccionar la carpeta de nombre *RapidMiner 5*, y dar click sobre la aplicación que deseamos iniciar.

7.3.3 Trabajando con RapidMiner

Una vez iniciada la aplicación es posible seleccionar si se desea crear un nuevo archivo, abrir un archivo reciente, abrir otros archivos, abrir una plantilla o ingresar a un tutorial en línea, tal y como se muestra en la figura 7.40.



Figura 7.40 Pantalla de Inicio de RapidMiner

Primeramente se seleccionará la opción *New*, para lo cual daremos click sobre la opción deseada. Una vez seleccionada esta opción, se requiere que seleccionemos una ubicación para el repositorio de datos; si se despliega el menú de la opción de *NewLocalRepository* aparece la base de datos creada para RapidMiner durante su instalación, en este caso es tesis y se seleccionará dicha opción como se muestra en la figura 7.41. Después de seleccionada se da click en *ok*.

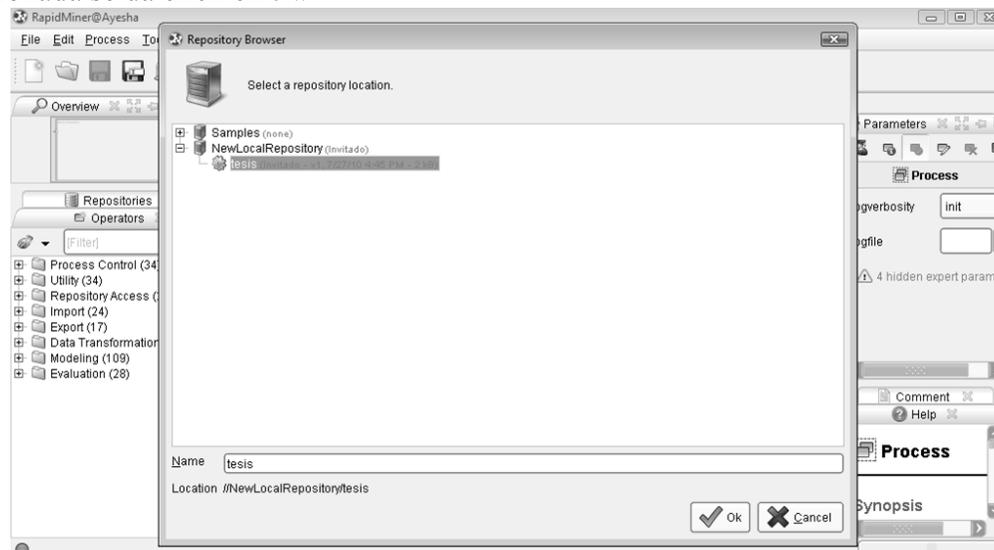


Figura 7.41 Selección del Repositorio de Datos

A continuación RapidMiner muestra un área en blanco en el centro de la pantalla, la cual será el área de trabajo; en este espacio se realizan diagramas de bloques; estos representan un algoritmo o proceso dentro del programa y son seleccionados del menú desplegable que se encuentra a la izquierda del espacio de trabajo. Si se desea que RapidMiner lea datos de un archivo de Excel, se debe seleccionar del menú desplegable la opción *Import*, después

Data, y por último *Read Excel*; en esta última opción se puede dar doble click sobre la selección o se puede arrastrar hasta el área de trabajo, como se muestra en la figura 7.42.

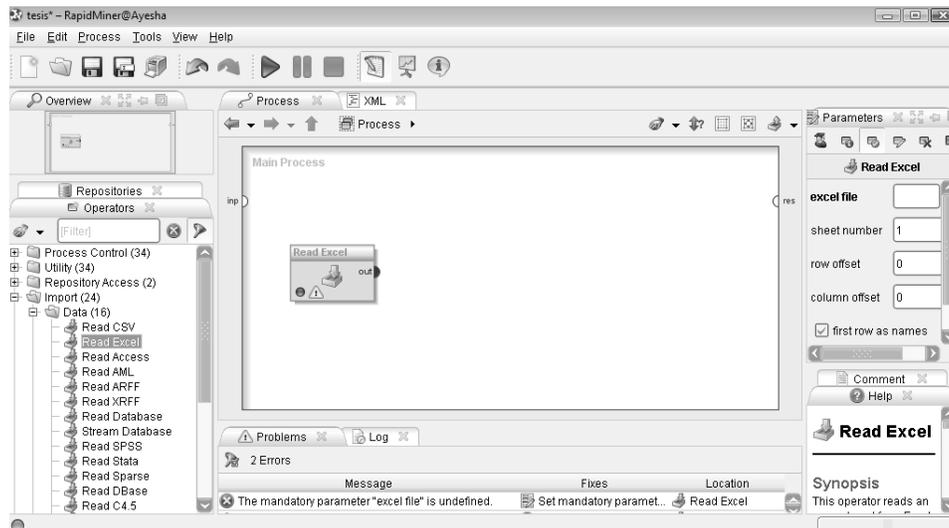


Figura 7. 42 Proceso de Importación de Datos

Se puede observar que dando click sobre el elemento anterior, a la derecha del área de trabajo se muestran las características del elemento, en este caso se pide que se seleccione el archivo de Excel que se desea que lea; en la parte inferior central de la ventana de RapidMiner se puede observar que tiene un área en la que se muestran los errores en el diagrama, solución y el lugar del diagrama donde es originado este problema.

Si lo que se desea hacer es que se realice el algoritmo *Random Forest* sobre los datos importados, se debe de seleccionar del menú desplegable el algoritmo *Random Forest* y arrastrarlo hacia el área de trabajo; una vez que se cuenta con el bloque de importación de datos y del algoritmo *Random Forest*, se deben de conectar ambos bloques y sus salidas. Esto es posible realizarlo al colocar el puntero del mouse sobre la salida del bloque de datos y arrastrar el cable hasta la entrada del bloque del algoritmo *Random Forest*; de la misma forma se conectan las salidas del bloque del algoritmo hacia el conector *res*.

Para poner en ejecución el proceso, se tiene que presionar con el mouse el botón de *play* que se encuentra en la parte superior de la ventana de RapidMiner. Una vez ejecutado el procedimiento, RapidMiner pregunta al usuario se desea ver en ese momento el resultado de la ejecución, al presionar en *yes* se observan los resultados obtenidos después de la ejecución.

Si durante el proceso de conexión de los bloques, el programa encuentra un error de los datos o de conexiones, Rapid Miner indica al usuario en el área de mensajes, lo que debe de realizar para corregir el error; es posible dar doble click sobre el mensaje, y Rapid Miner realiza la corrección automáticamente.

El proceso anterior se puede realizar para cualquier tarea de minería de datos que se desee ejecutar mediante este software; cabe señalar, que Rapid Miner es sólo una herramienta y

que el usuario define con su criterio, la validez de los resultados obtenidos utilizando este software.

Si lo que se desea obtener es información sobre una división, departamento o asignatura, desde una base de datos, se procede a realizar la consulta desde OpenDB sobre la base de datos; los primeros pasos a seguir para realizar una consulta de este tipo son:

- Descargar desde la página de Oracle el software ojdbc.jar correspondiente a la versión de Oracle que se está manejando.
- Instalar el ojdbc.jar dentro de la carpeta de WEKA.
- Iniciar WEKA desde modo consola o MS-DOS mediante el siguiente comando:
javaw -Xmx512m -classpath .;weka.jar;c:\ojdbc14.jar weka.gui.Main
- Seleccionar modo Explorer de WEKA.
- Conectarse a la base de datos desde la ventana de OpenDB (pp.44).
- Una vez conectado, realizar la consulta o query desde la ventana de Open DB.

Una vez realizados los pasos anteriores, debemos seleccionar el método que se desea llevar a cabo desde las pestañas que se encuentran en la parte superior de la ventana de WEKA.

Si el usuario desea llevar a cabo un análisis de la asignatura a su cargo, deberá ingresar la información correspondiente a las encuestas de todos los grupos de la asignatura ya sea de forma manual, a través de la base de datos o por medio de un archivo de Excel, y seleccionar un método de los vistos anteriormente, para que, de esta forma pueda obtener información adicional sobre el comportamiento de su asignatura, departamento o división a lo largo del tiempo.

CONCLUSIONES

- En el capítulo uno se presentan los conceptos necesarios para entender qué es y para qué sirve la minería de datos, las aplicaciones en las que se puede utilizar, así como los tipos de datos con los que se puede trabajar. Se considera a la minería de datos una herramienta muy útil que proporciona la información (nuevo conocimiento útil y novedoso) que se necesita para apoyar a la toma de decisiones.
- En el capítulo dos se describen los tipos de datos con los que se puede trabajar o que son válidos tanto para llevar a cabo un proceso de Minería de Datos como para el manejo de información en Oracle; también se introdujo a algunos procesos de manejo de información como la reducción de la dimensionalidad, entre otros.
- En el capítulo tres se describen ampliamente los pasos para la Minería de Datos, así como se menciona qué algoritmos son utilizados para cada etapa y para qué métodos.
- En el capítulo cuatro se presenta el software libre disponible para Minería de Datos, su manejo indispensable y alguna de sus características principales.
- En el capítulo cinco se presenta un análisis de la encuesta elaborado por el Ing. José Enrique Larios Canale, profesor de carrera de la Facultad de Ingeniería, en el que se detalla las preguntas que son pertinentes o no al alumno y porqué.
- En el capítulo seis se presenta un análisis de la encuesta previo a la Minería de Datos, en el que se detalla cómo se encuentra dividida la encuesta, cuántas secciones tiene y los elementos que se obtienen como resultados de la evaluación del desempeño del profesor.
- En el capítulo siete se lleva a cabo el desarrollo de Minería de Datos sobre la encuesta de profesores, a través de una serie de algoritmos y tareas descriptivas de las cuales se obtuvieron una serie de resultados y conclusiones. Finalmente se realizó un manual de usuario para que esta tesis pueda servir como elemento de apoyo en el proceso de toma de decisiones y se pueda continuar realizando Minería de Datos sobre las encuestas de profesores en la Facultad de Ingeniería.

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 podemos 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 o obtuvo NP 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.

La encuesta para profesores de la Facultad de Ingeniería, es un instrumento que tiene tanto errores como bondades, es decir, este instrumento no es perfecto, pero debido al tiempo que éste lleva aplicándose, constituyen una fuente de información importante, por lo cual debe de ser aprovechada, evaluando de antemano, la asertividad de algunas de sus preguntas a si como la interpretación que el alumno puede dar de ellas y que implica que en algunas de las preguntas se pueda dar un resultado subjetivo, dependiente del panorama que el alumno tiene sobre su profesor; de lo anterior se deriva la importancia que tiene el número de encuestados en el resultado final del profesor, ya que el punto de vista de 3 alumnos no es igual de significativo, que el punto de vista de 50 alumnos.

Sin embargo, la minería de datos trabajaba con cifras, cantidades y valores estadísticos, por lo que convierte a la interpretación humana un valor de suma importancia para el resultado final de este proceso, ya que el usuario le da el sentido adecuado a los resultados y orienta sobre su uso, fiabilidad y consistencia.

Finalmente, en esta tesis se propone que sean tomadas en cuenta para realizar Minería de Datos con esta herramienta, sólo algunas preguntas que reflejan de mejor forma la opinión del alumno sobre el profesor.

BIBLIOGRAFÍA

[1]Data Warehousing “La integración de la información para la mejor toma de decisiones”, Harjinder S. Gill, Praskash C. Rao. México, D.F., Prentice Hall Hispanoamérica, 1996.

[2]“Introduction to Data Mining”, Pang-Ning Tan, Steinbach, Kumar . México, D.F, Pearson Addison Wesley, 2006.

[3]Willi Klösgen. Handbook of Data Mining and Knowledge Discovery. Oxford University, 1ra edición, Oxford 2002.

[4]José Hernández Orallo, Ma. José Ramírez Quintana, César Ferri Ramírez. Introducción a la Minería de Datos. Pearson – Prentice Hall. Madrid, 2004. 1ra edición

[5]David Hand, Heikki Mannila, Padhraic Smyth.”Principles of Data Mining”. MIT Press, 1ra edición, USA, 2001.

[6]Ian H. Witten, Eibe Frank. Data Mining. Practical Machine Learning tools and Techniques with Java Implementations. Morgan Kaufmann Publishers. 1ra edición, USA, 2000.

[7]Han Kamber, Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, Primera edición, USA, 2001.

[8]Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka. Elements of Artificial Neural Networks. The MIT Press. 1ra edición, Cambridge, Massachusetts, 1997.

[9]Regresión: <http://www.bioestadistica.uma.es/libro/node40.htm>

[10]Rapidminer: <http://rapid-i.com>. Sitio web del software de minería de datos. Última consulta en Mayo del 2010.

[11]SPSS : <http://www.spss.com/>. Sitio web del paquete de análisis estadístico. Última consulta en Mayo del 2010.

[12]Base de Datos otorgada por la Unidad de Servicios de Cómputo Administrativos a través de la Secretaría de Servicios Académicos de la Facultad de Ingeniería.