

**Universidad de las Ciencias Informáticas**



Trabajo de Diploma para optar por el Título de  
Ingeniero en Ciencias Informáticas

Informatización de los diagnósticos realizados a los estudiantes de primer año de la facultad 5 de la Universidad de las Ciencias Informáticas en el análisis del rediseño.

Autor: Orlys Valero Tamayo

Tutor: Msc. Roberto Millet Luaces

La Habana Junio, 2011

## Declaración de Autoría

---

Declaro ser la única autora del trabajo “Informatización de los diagnósticos realizados a los estudiantes de primer año de la facultad 5 de la Universidad de las Ciencias Informáticas en el análisis del rediseño.” y autorizo a la Facultad 5 y a la Universidad de Ciencias Informáticas a hacer uso del mismo, con carácter exclusivo.

Para que así conste firmo la presente a los \_\_\_\_\_ días del mes \_\_\_\_\_ del año \_\_\_\_\_.

Autora:

---

Orlys Valero Tamayo

Tutor:

---

Msc. Roberto Millet Luaces



## Datos de Contacto

---

Datos de Contacto:

**Tutor:** Roberto Millet Luaces.

**Breve currícul:**

- Profesor de Matemática.
- Graduado de Ingeniero Eléctrico en 1986, en Universidad de Camagüey.
- Profesor Auxiliar.
- Msc en Ciencias Matemáticas.
- Imparte docencia en universidades desde 1987.

**Ubicación:** UCI, Cuba.

**E-mail:** [milletp@uci.cu](mailto:milletp@uci.cu)



## Agradecimientos

---

### *Agradecimientos:*

*A mi mamá Marlene y a mi papá Antonio Marcial por su apoyo, amor y confianza en todas las decisiones que he tomado en la vida.*

*A mi tutor Millet, por su guía, apoyo y confianza.*

*A mi hermana por ser tan linda y tan buena conmigo.*

*A mi papá Fernando y a mi mamá Fe por todo su cariño, ayuda, apoyo y confianza en su nieta.*

*A Kadier por robarse mi corazón y hacer de mí la chica más feliz de mundo, por su apoyo, amor y confianza para que yo realizara mis sueños.*

*A mis tíos y primos Mayila, Maribel, Fernando, Yusbel, Bernardo, Yosdaivys, Katia y en general a toda la familia que no cabe en esta página.*

*A mi amiga Nora aunque no pueda estar conmigo para disfrutar juntas mi alegría.*

*A mis compañeros Riolvis, Alejandro y Gabriel por estar siempre que los necesite.*

*A mis amigos Kimbo, Pedro, Neysis, Silvia, Reineris, Frank, Lila, Claudia, Yoneida, Ana Leonor, Idaisy, Marelis, Odalis e Hildita por su ayuda y preocupación en todos estos años.*

*A mis compañeras del equipo de fútbol y a mi entrenador, padre y amigo Ariel Alfonso More.*

*A mis compañeras de apartamento Sulemy, Eileen, Danae, Gabriel por acogerme con tanta alegría y cariño.*

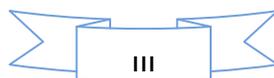
*A todos mis maestros y profesores por enseñarme lo que sé.*

*A la revolución por esta escuela tan linda.*

*A Olga Lidia, Walfrido y al CICE por facilitarme los datos que necesitaba.*

*A todo aquel que ocupó, ocupa u ocupará un lugar en mi corazón.*

Orlys Valero Tamayo.



## Dedicatoria

---

*Dedicatoria:*

*A mi mamá y a mi papá como un regalo.*

*A mi abuelo como el regalo más grande que nunca ha podido tener.*

*A mi novio Kadier como uno de tantos regalos que me faltan por darle en la vida.*

*A mi hermana para que siga como va y me tenga como ejemplo.*

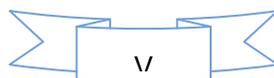
### RESUMEN

La Inteligencia Artificial (IA) en la actualidad está en constante desarrollo, con aplicaciones en diversos sectores. Dentro de la IA se han desarrollado de forma vertiginosa las técnicas que permiten adquirir conocimiento de grandes conjuntos de datos. La teoría de los conjuntos aproximados es una de las ramas de la IA relativamente nueva, que permite procesar grandes cantidades de variables que intervienen en un problema determinado, además proporciona un conjunto de herramientas potentes para la resolución de problemas complejos.

El presente trabajo de diploma surge debido a la necesidad que existe de buscar nuevos métodos de enseñanza que contribuyan a mejorar el proceso de enseñanza-aprendizaje de los estudiantes de la Facultad 5 de la Universidad de las Ciencias Informáticas (UCI), a partir de los datos obtenidos en el diagnóstico aplicado al inicio del curso.

Por este motivo, se decidió desarrollar una aplicación que permita conocer las características cognoscitivas de los estudiantes que se matriculan en la Facultad 5 de UCI.

**Palabras Clave:** *inteligencia artificial, minería de datos, teoría de conjuntos aproximados.*



# Índice de Contenido

---

Introducción .....	1
Capítulo 1: Fundamentación Teórica. ....	4
Introducción al capítulo.....	4
1.1 Inteligencia Artificial.....	4
1.2 Conceptos relacionados con Inteligencia Artificial. ....	6
1.2.1 Incertidumbre .....	6
1.2.2 Aprendizaje y Conocimiento.....	6
1.2.3 Aprendizaje.....	6
1.3 Teoría de los Conjuntos Aproximados. ....	7
1.3.1 Representación de la RST. Sistema de Información. ....	8
1.3.2 Relación de separabilidad. Matriz de separabilidad.....	9
1.3.3 Propiedades de la RST .....	10
1.4 Minería de Datos. ....	11
1.5 Selección de Rasgos. ....	12
1.5.1 Selección de Atributos.....	13
1.6 Teoría de Grafo. ....	15
1.6.1 Árbol de decisión.....	16
1.7 Matriz Booleana.....	17
1.8 Lógica Difusa.....	17
1.9 Descripción de las herramientas y técnicas utilizadas. ....	18
1.9.1 Software Weka.....	18
1.9.1.1 Herramientas y Ventajas del software Weka.....	19
1.9.2 Statgraphics. ....	19
1.9.3 Lenguaje de programación Qt/C++. ....	20
1.9.4 Método Delphi de la consulta de expertos.....	22
Capítulo 2: Aplicación del conocimiento. ....	24
Introducción al capítulo. ....	24
2.1 Método Delphi de la consulta de expertos. ....	24
2.1.1 Encuesta aplicada a los expertos.....	24
2.1.2 Resultados obtenidos.....	26
2.1.3 Análisis estadístico de la información.....	31
2.1.4 Gráficos obtenidos. ....	34
2.2 Proceso de descubrimiento del conocimiento. Análisis de los estudiantes encuestados. 36	
2.2.1 Análisis de los Diagnósticos.....	36

## Índice de Contenido

---

2.2.2 Base de Hechos.....	48
2.3 Procesamiento de datos.....	50
2.3.1 Preparación de los datos en el WEKA.....	50
2.3.2 Procesamiento de la Base de Hechos y resultados obtenidos. ....	51
2.4 Árboles de decisión. ....	52
Capítulo 3: Resultados Obtenidos.....	54
Introducción al capítulo .....	54
3.1 Sistema de Reglas. ....	54
3.1.1 Selección del clasificador de tipo regla.....	54
3.1.2 Reglas.....	55
3.2 Solución al problema planteado.....	60
3.2.1 Características de la aplicación.....	60
3.2.2 Interfaces de la Aplicación. ....	60
Conclusiones .....	64
Recomendaciones .....	65
Referencias bibliográficas .....	66
Bibliografía.....	68

### **Introducción**

La UCI para cumplir con su misión estratégica de formar profesionales altamente calificados en la producción de software y servicios informáticos, requiere profundos cambios en su modelo de formación estudio-producción y centra su atención en el proceso de aprendizaje del estudiante. Por estos motivos al inicio del curso en la UCI se realiza un diagnóstico que consiste en la aplicación de 7 ejercicios: estilo decisional, estrategia de aprendizaje, perfil de inteligencia, lengua materna, idioma inglés, matemática y cultura económica.

Como se puede observar los temas a tratar en cada ejercicio tienen contenidos diversos y en correspondencia a las respuestas dadas por los estudiantes se hace necesario tomar decisiones que estarán en función de encontrar aptitudes cognoscitivas y características de los mismos, que constituyan a la búsqueda de métodos de enseñanza que ayuden a la formación integral de los educandos.

Actualmente el Centro de Innovación y Calidad de la Educación (CICE) de la UCI, no cuenta con una aplicación capaz de procesar los resultados obtenidos en cada uno de los ejercicios, solo dispone de un sitio donde son publicados los 7 cuestionarios. Luego que el estudiante responde los mismos y envía sus respuestas, este guarda los resultados para procesarlos y arribar a un criterio de cada estudiante, el criterio obtenido no es completamente fiable porque los encargados del análisis de las encuestas utilizan el método de apreciación que no es el óptimo para arribar a conclusiones y es necesario que las respuestas dadas sean completamente fiable por lo que se debe utilizar el método psicológico o el matemático, además tampoco se tiene en cuenta la procedencia estudiantil ni los datos personales del estudiante, siendo estos factores importantes que influyen en la educación de los mismos.

Con el propósito de dar respuesta a las interrogantes planteadas, a continuación se presenta el perfil del trabajo de diploma.

#### Problema a Resolver

¿Cómo obtener las principales características cognoscitivas de los estudiantes de primer año de la facultad 5 mediante los diagnósticos aplicados al inicio del curso?

# Introducción

---

## Objeto de Estudio

Proceso de obtención de los principales rasgos para la caracterización de los estudiantes en la aplicación de un nuevo modelo de enseñanza.

## Objetivo General

Desarrollar una aplicación que permita determinar las principales características cognoscitivas de los estudiantes que ingresan en el 1er año en la facultad 5 de la UCI.

## Campo de Acción

Aplicación de la teoría de conjuntos aproximados en la obtención de los principales rasgos de los estudiantes de 1er año de la facultad 5.

## Tareas Investigativas

- Consulta de la bibliografía del tema de investigación para el análisis del estado del arte.
- Estudio de la teoría de los conjuntos aproximados para aplicarla a la investigación.
- Análisis de los diagnósticos aplicados a los estudiantes, para confeccionar el sistema de información.
- Análisis del método Delphi para validar la necesidad e importancia del tema de investigación.
- Estudio del asistente estadístico Statgraphics para procesar la información obtenida.
- Estudio del software WEKA, para la obtención de las reglas de producción.
- Desarrollo de una aplicación que permita determinar las principales características cognoscitivas de los estudiantes.

## Idea a Defender

Si se aplica la teoría de los conjuntos aproximados en los resultados de los diagnósticos realizados a los estudiantes de primer año de la facultad 5, entonces se obtendrá una aplicación que permitirá determinar sus principales características cognoscitivas.

# Capítulo 1: Fundamentación Teórica

---

## Métodos científicos

### **Métodos Teóricos:**

- Analítico-sintético: Se utilizó para el desglose de la información en las diferentes áreas de importancia según los objetivos de la investigación, para facilitar el estudio de la inteligencia artificial, la reducción de atributos, la teoría de los conjuntos aproximados entre otras. Luego se integra la información obtenida para conformar el estado del arte y otros aspectos básicos de la investigación.
- Análisis histórico lógico: Se utilizó para el estudio de la evolución y desarrollo histórico de la teoría de conjuntos aproximados y otras disciplinas como la inteligencia artificial y la minería de datos. Además para caracterizar la evolución de la funcionalidad de los métodos existentes.
- Inductivo deductivo: Se utilizó para obtener reglas de inferencia a partir de datos y regularidades iniciales para llegar a conclusiones y viceversa.

### **Métodos Empíricos:**

- Encuesta: Se utilizó en la consulta de expertos.
- Consulta de expertos: Se utilizó para obtener el criterio de los especialistas.

## Capítulo 1: Fundamentación Teórica.

### Introducción al capítulo.

El capítulo contendrá todo el contenido teórico del presente trabajo. En él se describen los principales conceptos como resultado de la investigación realizada para la aplicación de la teoría de conjuntos aproximados en la obtención de los principales rasgos de los estudiantes de primer año de la facultad 5 para la implementación del modelo de rediseño en la docencia. Se ha tomado como muestra el 70% de las encuestas realizadas a primer año lo que representan 784 respuestas procesadas para la investigación; además se determinarán cuáles son las herramientas que se utilizarán.

Para procesar esta gran cantidad de datos y utilizarlos eficientemente, se hace necesaria la presencia de una tecnología de apoyo, para explorar, analizar, comprender y aplicar el conocimiento obtenido utilizando grandes volúmenes de datos.

### 1.1 Inteligencia Artificial.

La IA es la rama de la informática que desarrolla procesos que imitan a la inteligencia de los seres vivos. La principal aplicación de esta ciencia es la creación de máquinas para la automatización de tareas que requieran un comportamiento inteligente. (1)

Algunos ejemplos se encuentran en el área de control de sistemas, planificación automática, la habilidad de responder a diagnósticos y a consultas de los consumidores, reconocimiento de escritura, reconocimiento del habla y reconocimiento de patrones. Los sistemas de IA actualmente son parte de la rutina en campos como economía, medicina, ingeniería y la milicia, y se ha utilizado en gran variedad de aplicaciones de software, juegos de estrategia como ajedrez de computador y otros videojuegos.

A continuación se describe la evolución de los conceptos de IA desde el punto de vista de diferentes autores:

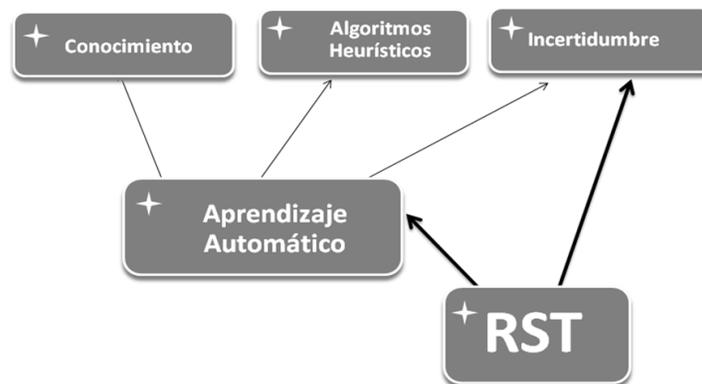
- “La automatización de actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisión, resolución de problemas, aprendizaje...” Bellman, 1978.

## Capítulo 1: Fundamentación Teórica

---

- “El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mente, en el más amplio sentido literal”. Haugeland, 1985. “El estudio de las facultades mentales mediante el uso de modelos computacionales”.
- Charniak y McDermott, 1985. “La Inteligencia Computacional es el estudio del diseño de agentes inteligentes”. Pool y col. 1988.
- “El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia”. Kurzweill, 1990.
- “El estudio de cómo lograr que los computadores realicen tareas que, por el momento, las personas hacen mejor”. Rich y Knight, 1991.
- “El estudio de los cálculos que hace posible percibir, razonar y actuar”. Winston, 1992.
- “La capacidad que tienen las máquinas para realizar tareas que en el momento son realizadas por seres humanos”. Rich & Knight 1994, Stuart 1996.
- “La IA... está relacionada con conductas inteligentes en artefactos”. Nilsson, 1998. (1)

Existen tres componentes principales en el estudio de la IA, el conocimiento, los algoritmos heurísticos y la incertidumbre. Luego aparece otro campo denominado Aprendizaje Automático y más tarde la Teoría de Conjuntos Aproximados (RST) la cual crea una mezcla entre el aprendizaje automático y la incertidumbre. A continuación se muestra la dependencia existente entre estas ramas del saber.



**Figura 1: Relación Aprendizaje Automático, RST, Conocimiento, Algoritmos Heurísticos e Incertidumbre.**

### 1.2 Conceptos relacionados con Inteligencia Artificial.

#### 1.2.1 Incertidumbre

La palabra incertidumbre significa duda o indecisión. Situación en la cual no se conoce completamente la probabilidad de que ocurra un determinado evento. Puede derivarse de una falta de información o incluso por que exista desacuerdo sobre lo que se sabe o lo que podría saberse. Puede tener varios tipos de origen, desde errores cuantificables en los datos hasta terminología definida de forma ambigua o previsiones inciertas del comportamiento humano .La incertidumbre puede manifestarse de diversas formas y ser provocada por diferentes causas, pero a grandes rasgos se pueden reconocer tres grandes categorías de incertidumbre:

- Provocada por información o conocimiento impreciso.
- Provocada por información incompleta.
- Provocada por conceptos o palabras inexactas.

#### 1.2.2 Aprendizaje y Conocimiento.

Al hablar de obtención de conocimiento, se está hablando de diversas técnicas de inducción del mismo, surgiendo entonces la pregunta ¿Qué es el conocimiento?

**Conocimiento:** Es el entendimiento sobre un determinado tema, en el cual una persona pueda expresar su opinión. En el caso de la IA este concepto podría variarse un poco, siendo el conocimiento conjunto de patrones que se obtienen de grupos de datos. También existe otro concepto que está estrechamente relacionado con el anterior.

#### 1.2.3 Aprendizaje

No es más que adquirir conocimiento sobre alguna cosa o tema. Existen distintos tipos de aprendizaje a través de los cuales se puede obtener conocimiento:

- **Aprendizaje Inductivo:** se pretende crear modelos de conceptos a partir de la generalización de conjuntos de ejemplos.
- **Aprendizaje Analítico o Deductivo:** se aplica la deducción para obtener descripciones generales a partir de un ejemplo de concepto y su explicación.

- **Aprendizaje por Refuerzo:** el algoritmo aprende observando el mundo que lo rodea.
- **Aprendizaje Automático o Aprendizaje de Máquinas:** es una rama de la IA que se ocupa del procesamiento de datos y del desarrollo de técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos. (2)

Para el desarrollo de casi todos los algoritmos relacionados con el aprendizaje se utilizan las Bases de Hechos que son creadas a partir de conjunto de ejemplos, los cuales tienen que ver con el tema que se quiera desarrollar, por tanto una Base de Hechos es una estructura de datos que contiene información sobre una situación concreta.

### 1.3 Teoría de los Conjuntos Aproximados.

Una herramienta matemática muy potente para la selección, es la teoría de conjuntos aproximados (Rough Sets Theory), propuesta por el profesor polaco Z. Pawlak y su equipo en 1982. (3) La filosofía de conjuntos aproximados se basa en asumir que existe información asociada con cada objeto del universo. (4) Un conjunto de entrenamiento se representa por una tabla donde cada fila representa un objeto y cada columna un atributo, a este conjunto se le llama sistema de información, más formalmente, es un par  $S = (U, A)$  donde  $U$  es un conjunto no vacío y finito de objetos llamado universo y  $A$  es un conjunto no vacío y finito de atributos. Un sistema de decisión es cualquier sistema de información de la forma  $SD = (U, A \cup \{d\})$  donde  $d \in A$  es un atributo de decisión, o sea, un atributo que indica a qué grupo pertenece el objeto.

El modelo de los conjuntos aproximados posee importantes ventajas en el análisis de datos. La principal se basa únicamente en los datos originales y no requiere de información externa para obtener conocimiento sobre el sistema, de forma que no es necesario hacer suposiciones sobre un nuevo algoritmo de selección de rasgos; la otra ventaja importante consiste en que esta herramienta permite analizar atributos tanto cuantitativos como cualitativos. Según la universidad de Kentucky de los EE.UU. RST “Es una importante herramienta en el análisis de datos con importantes aplicaciones en la minería de datos y el descubrimiento del conocimiento”. (5)

### 1.3.1 Representación de la RST. Sistema de Información.

La RST se representa de la siguiente forma:

RST= Sistema de Información + Relación de Separabilidad.

Donde, un sistema de información no es más que los datos del dominio, y la relación de separabilidad está constituida por las componentes de la teoría. Si se tiene en cuenta que a todo objeto se le puede asociar algún tipo de información basados en sus atributos/características, entonces es factible representar estos atributos por medio de una tabla o sistema de información. Dentro de éste sistema de información, las filas representan los objetos, en cuanto que las columnas representan los atributos. Las entradas de la tabla, que no son otra cosa, más que pares (objeto, atributo), vienen a ser los valores de cada objeto para cada atributo. La tabla de información debe ser entendida, en términos prácticos, como una matriz finita, tal como se observa en la tabla 1.

Objeto m	Valor m, 1	Valor m, 2	.....	Valor m, n
	Atributo 1	Atributo 2	.....	Atributo n
Objeto 1	Valor 1, 1	Valor 1, 2	.....	Valor 1, n
Objeto 2	Valor 2, 1	Valor 2, 2	.....	Valor 2, n
.....	.....	.....	.....	.....

**Tabla 1: Representación de un sistema de información.**

Sea un conjunto de atributos  $A = \{A_1, A_2, \dots, A_n\}$  y un conjunto  $U$  de ejemplos no vacío llamado universo (objetos, entidades, etc.) descritos usando los atributos  $A_i$ . Un sistema de información se convierte en un sistema de decisión cuando a cada elemento de  $U$  se le agrega un nuevo atributo "d" llamado decisión, entonces sistema de decisión sería:  $(U, A \cup \{d\})$ :  $d \notin A$  donde  $d$  es el rasgo de decisión.

La tabla 2 muestra algunos elementos de decisión donde cada fila representa un objeto y cada columna un atributo.

	Materias Primas y Materiales	Portadores Energéticos	Servicios inmobiliarios y de Protección	D
Software 1	3251-4375	1951-2825	60-149	NO
Software 2	5501-6625	2826-3700	326-413	YES
Software 3	8876-10000	6326-7200	414-500	YES
Software 4	7751-8875	4576-5450	238-325	YES
Software 5	1000-2125	200-1075	150-237	NO
Software N	4376-5500	1076-1950	60-149	YES

**Tabla 2: Representación de un sistema de decisión. (6)**

### 1.3.2 Relación de separabilidad. Matriz de separabilidad.

Un atributo  $A_i \notin A$ , siendo  $A$  el conjunto de atributos inicial, separa un objeto 'X' de un objeto 'Y' si y sólo si los valores para el mismo atributo en elementos diferentes sean totalmente diferentes.

$$(R_1): \text{Separa}(A_i, x, y) \Leftrightarrow f(x, A_i) \neq f(y, A_i).$$

Una matriz de separación (MA) es una matriz  $|U|_x|U|$  donde cada entrada  $MA(x, y) \subseteq A$  contiene el conjunto de atributos que distinguen los elementos  $x$  e  $y$  de  $U$ .

$$MA(x, y) = \{A_i \in A : \text{Separa}(A_i, x, y)\}$$

### 1.3.3 Propiedades de la RST

La teoría de conjuntos aproximados tiene un conjunto de propiedades que se muestran a continuación:

La aproximación inferior del conjunto  $X$   $B_i(X)$  es un subconjunto del propio conjunto  $X$  y a su vez, este es un subconjunto de la aproximación superior de  $X$   $B_s(X)$   $B_i(X) \subseteq X \subseteq B_s(X)$

La aproximación inferior del conjunto vacío  $B_i(\{\})$  es igual a la aproximación superior del conjunto vacío  $B_s(\{\})$ , siendo estos igual al propio conjunto vacío.  $B_i(\{\}) = B_s(\{\}) = \{\}$

La aproximación inferior del universo  $B_i(U)$  es igual a la aproximación superior del universo  $B_s(U)$ , siendo estos igual al propio universo  $B_i(U) = B_s(U) = U$

La aproximación superior de la unión de dos subconjuntos  $X$  e  $Y$  del universo  $B_s(X \cup Y)$  es igual a la aproximación superior del subconjunto  $X$  unido con la aproximación superior del subconjunto  $Y$   $B_s(X) \cup B_s(Y)$ .

$$B_s(X \cup Y) = B_s(X) \cup B_s(Y)$$

La aproximación superior de la intersección de dos subconjuntos  $X$  e  $Y$  del universo  $B_s(X \cap Y)$  es igual a la aproximación inferior del subconjunto  $X$  interceptado con la aproximación inferior del subconjunto  $Y$   $B_i(X) \cap B_i(Y)$ .

$$B_s(X \cap Y) = B_i(X) \cap B_i(Y)$$

Un subconjunto  $X$  es subconjunto de otro subconjunto  $Y$ , siendo  $X$  e  $Y$  subconjuntos del universo, si y sólo si la aproximación inferior de  $X$   $B_i(X)$  es subconjunto de la aproximación inferior de  $Y$   $B_i(Y)$  la aproximación superior de  $X$   $B_s(X)$  es subconjunto de la aproximación superior de  $Y$   $B_s(Y)$ .

$$X \subseteq Y \Rightarrow B_i(X) \subseteq B_i(Y) \text{ y } B_s(X) \subseteq B_s(Y)$$

La aproximación inferior de un subconjunto  $X$  del universo  $U$   $B_i(X)$  es igual al universo  $U$  menos la aproximación superior del universo menos el conjunto  $X$   $B_s(U - X)$   
$$B_i(X) = U - B_s(U - X)$$

### 1.4 Minería de Datos.

La Minería de datos es el proceso de descubrir patrones de información interesante y potencialmente útiles, inmersos en una gran base de datos en la que se interactúa constantemente. (5) De forma general, los datos constituyen la materia prima, en el momento que el usuario les atribuye algún significado especial pasan a convertirse en información.

La minería de datos (data mining) surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos; compuesta por etapas que integra varias áreas y que no se debe confundir con un gran software. (7) Durante el desarrollo de un proyecto de este tipo se emplean diferentes aplicaciones de software en 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 facilitan el desarrollo de un proyecto, sin embargo, casi siempre acaban complementándose con otra herramienta. El verdadero trabajo de la minería de datos es reunir las ventajas de varias áreas como la Estadística, la Inteligencia Artificial, la Computación Gráfica, las Bases de Datos y el Procesamiento Masivo, utilizando como materia prima las bases de datos.

Una definición tradicional es la siguiente: "Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos". Desde otro punto de vista se define como "la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión (7).

La Minería de datos es una combinación de procesos como:

- Extracción de datos
- Limpieza de datos.
- Selección de características.
- Algoritmos.
- Análisis de resultados.

### **1.5 Selección de Rasgos.**

El proceso de selección de rasgos consta de dos componentes principales: una función de evaluación y un método de búsqueda. La función de evaluación permite calcular la calidad de un subconjunto de rasgos; mientras que el método de búsqueda, por lo general heurística, es el encargo de generar los subconjuntos de rasgos; dados  $N$  rasgos se tienen  $2^N - 1$  posibles subconjuntos (8).

Grandes volúmenes de datos se pueden manipular rápidamente al extraer de ellos sólo la información necesaria que los describa, sin perder la calidad del sistema, y obteniendo conocimiento sobre ellos. Los conjuntos de datos consisten en un sistema de información descrito por un conjunto de atributos y los objetos que se representan como combinaciones de valores válidos dentro del dominio de cada atributo. De esta forma, la selección de rasgos en un sistema de información de este tipo, consiste en obtener un subconjunto de atributos tal que describa el sistema como si se tratara del conjunto completo. Esto quiere decir que el proceso se centra en encontrar los atributos más importantes dentro de los que se han utilizado para representar los datos y eliminar aquellos que se consideran irrelevantes y hacen más difícil el proceso de descubrimiento de conocimiento dentro de una base de datos.

### 1.5.1 Selección de Atributos.

La selección de atributos o características, como su nombre lo indica, selecciona un subconjunto de los atributos originales. Se puede considerar como un problema de búsqueda. (Ver Figura 2). Pretende elegir atributos que sean relevantes para una aplicación y lograr el máximo rendimiento con el mínimo esfuerzo. Permite mejorar la precisión e interpretación de los métodos de aprendizaje automático, además de reducir el tamaño de la base de datos y el tiempo de los algoritmos de aprendizaje. (9)

El resultado de la selección de atributos está dado por:

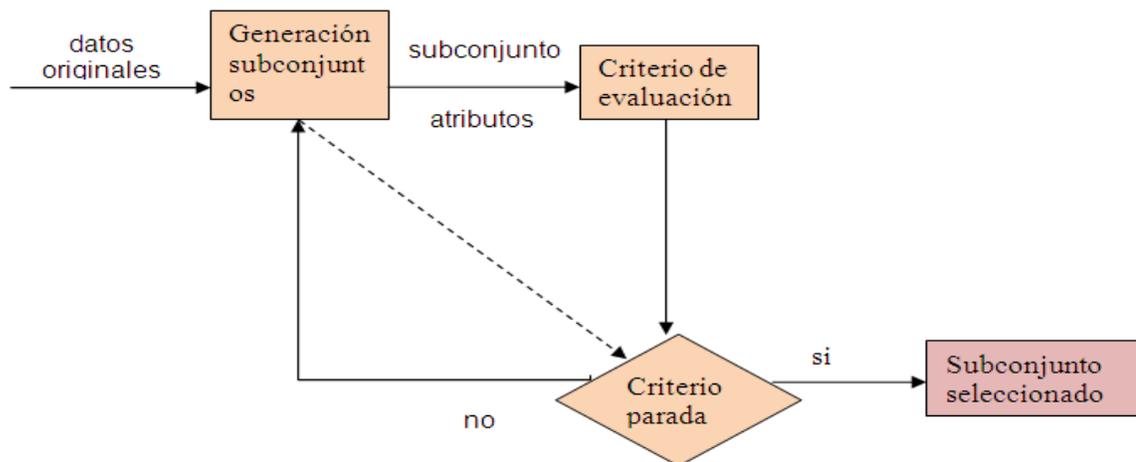
- Menos datos: los algoritmos pueden aprender más rápido.
- Mayor exactitud: el clasificador generaliza mejor.
- Resultados más simples: más fácil de entender.
- Menos atributos: evitar obtenerlos posteriormente.

Las ventajas esperadas de este proceso son:

- Mejorar el desempeño predictivo.
- Construir modelos eficientemente.
- Mejorar el entendimiento sobre los modelos generados.

El proceso de selección de atributos involucra cuatro pasos:

- Generación de candidatos: involucra una estrategia de búsqueda.
- Evaluación de candidatos: requiere un criterio de evaluación.
- Criterio de parada: este puede darse por la estrategia de búsqueda, el número de iteraciones realizadas, el número de atributos seleccionados, que no se mejore el criterio de evaluación al añadir o quitar otro atributo, que el error de clasificación está por debajo a un valor.
- Validación de resultados: si se conoce de entrada cuáles son los atributos relevantes, se puede comparar el resultado del algoritmo contra esos atributos conocidos. Como normalmente no se sabe, se puede comparar el error en la clasificación con y sin la selección de atributos.



**Figura 2: Proceso de la selección de atributos. (6)**

La selección de atributos tiene distintas clasificaciones, estas pueden ser:

Según la evaluación: filtros, envolturas, híbridos.

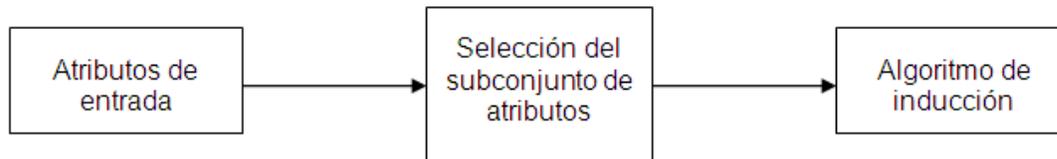
Disponibilidad de la clase: supervisados, no supervisado.

Según la búsqueda: completa, heurística, aleatoria.

Según la salida del algoritmo: lista ordenada de todos los atributos según su importancia, subconjunto de atributos.

En general, los algoritmos de selección de atributos se distinguen por su forma de evaluar atributos y se pueden clasificar en tres tipos:

- Filtros (del inglés filters): seleccionan los atributos en forma independiente del algoritmo de aprendizaje
- Figura 3: Esquema genérico de algoritmo tipo filtros.(10)
- Envolturas (del inglés wrappers): usan el desempeño de algún clasificador para determinar lo deseable de un subconjunto.
- Híbridos: usan una combinación de los dos criterios de evaluación en diferentes etapas del proceso de búsqueda. (7)



**Figura 3: Esquema genérico de algoritmo tipo filtros. (6)**

Cambiando la forma de generar nuevos candidatos y el criterio de parada, se pueden generar diferentes versiones.

- Diferentes criterios de evaluación generan diferentes algoritmos de tipo filtros.
- Diferentes algoritmos de aprendizaje generan diferentes algoritmos de tipo envoltura.

El otro tipo de algoritmo es el híbrido que combina el criterio de evaluación de ambos. La idea general del algoritmo es la siguiente:

- Empieza con un subconjunto  $S_0$ .
- Aumenta el tamaño del subconjunto inicial en un atributo (cardinalidad) y evalúa todos los elementos de esa cardinalidad con una medida independiente del algoritmo de aprendizaje.
- El mejor subconjunto se evalúa con un algoritmo de aprendizaje y se queda con el mejor (de todas las cardinalidades vistas hasta el momento).
- Continúa con la siguiente cardinalidad.
- En caso que no mejore la calidad del algoritmo de aprendizaje, esta condición se puede tomar como criterio de parada. (11)

### 1.6 Teoría de Grafo.

La teoría de grafo desempeña un papel importante en la fundamentación matemática de diversos problemas, en particular en la ciencia de la computación. Constituye una herramienta básica en la modelación de fenómenos discretos y es fundamental para comprensión de las

estructuras de datos y el análisis de algoritmos. Los grafos se conocen también como red, los mismos están compuestos por nodos y aristas y representan una generalización del concepto simple de un conjunto de puntos llamados vértices.

Un grafo  $G$  es un par  $G = (V, E)$ , donde  $V$  es un conjunto finito (vértices, nodos) y  $E$  es un multi-conjunto de pares no ordenados de vértices, denotados por  $\{X, Y\}$ , que se denominan lados, aristas, etc. En este caso  $X$  y  $Y$  son extremos de  $\{X, Y\}$ . Se denota  $V(G)$  como el conjunto de vértices del grafo  $G$  y por  $E(G)$  el conjunto de lados del grafo  $G$ . Además  $v(G)$  y  $e(G)$  denotan el número de vértices y el número de aristas de  $G$  respectivamente. Puesto que  $E$  es un multi-conjunto donde existen pares repetidos, en este caso  $G$  tiene lados múltiples. También es posible que algún par no ordenado de  $E$  tenga el mismo vértice repetido, en este caso se dice que el lado es un lazo (loop) o bucle. Cuando existen lados múltiples y/o lazos se dice que  $G$  es un multi-grafo. Si no hay lados múltiples ni lazos se dice que es un grafo simple.

### 1.6.1 Árbol de decisión.

Es un grafo explicando las secuencias de las decisiones a tomar y los diversos acontecimientos que pueden suceder. El árbol de decisión es una forma de representar la información que se dispone, reflejando las posibles alternativas en función de los sucesos inciertos que pueden acontecer. (12)

Elementos fundamentales en la elaboración de un árbol de decisión:

- Puntos de decisión.
- Acontecimientos.
- Resultados esperados.

En el árbol de decisión existen dos tipos de vértices que deben representarse con notación diferente:

- Puntos de decisión que representan las distintas opciones a adoptar por el decidor ante una situación dada.
- Sucesos inciertos que corresponden a los distintos fenómenos que pueden suceder cuando se ha adoptado una decisión.

Para elaborar un árbol de decisión se tienen en cuenta las siguientes etapas:

- Determinar el conjunto de acciones posibles y de acontecimientos posibles, así como la secuencia de las decisiones a tomar.
- Representar mediante un árbol las secuencias alternativas de acción o estrategias de acontecimientos.
- Partir del grafo, determinar los puntos finales. Para cada punto final se calculará el resultado condicional en función de los datos económicos disponibles.
- A cada acontecimiento se afectará entonces de una probabilidad. (12)

### 1.7 Matriz Booleana.

La Matriz Booleana es una matriz de decisión donde solo aparecen dos estados uno de ellos niega el otro, los elementos  $A_{i,j}$  toman valores 0 y 1. Su campo de aplicación es muy variado, se utiliza con frecuencia en las ramas de ingeniería y matemática, por ejemplo, en la consulta de experto tiene mucha aplicación cuando se realiza el análisis estadístico de las respuestas dadas por los expertos, en algebra lineal su aplicación es muy marcada, lo mismo ocurre en el análisis de los circuitos lógicos donde se hace necesario el análisis de dos estados.

### 1.8 Lógica Difusa.

La lógica difusa (borrosa) es una rama de la IA que se funda en el concepto "Todo es cuestión de grado", lo cual permite manejar información vaga o de difícil especificación, si se quisiera hacer cambiar con esta información el funcionamiento o el estado de un sistema específico, es entonces posible con la lógica borrosa gobernar un sistema por medio de reglas de 'sentido común' las cuales se refieren a cantidades indefinidas. (12)

Las reglas involucradas en un sistema borroso, pueden ser aprendidas con sistemas adaptativos que aprenden al 'observar' como operan las personas los dispositivos reales, o estas reglas pueden también ser formuladas por un experto humano. En general la lógica borrosa se aplica tanto a sistemas de control como para modelar cualquier sistema continuo de ingeniería, física, biología o economía.

La lógica borrosa es entonces definida como un sistema matemático que modela funciones no lineales, que convierte unas entradas en salidas acordes con los planteamientos lógicos que utilizan el razonamiento aproximado. Se fundamenta en los denominados conjuntos borrosos y

un sistema de inferencia borroso basado en reglas de la forma " SI..... ENTONCES..... ", donde los valores lingüísticos de la premisa y el consecuente están definidos por conjuntos borrosos, es así como las reglas siempre convierten un conjunto borroso en otro.

La mayoría de los fenómenos que se encuentran cada día son imprecisos, es decir, tienen implícito un cierto grado de difusidad en la descripción de su naturaleza. Esta imprecisión puede estar asociada con su forma, posición, momento, color, textura, o incluso en la semántica que describe lo que son. En muchos casos el mismo concepto puede tener diferentes grados de imprecisión en diferentes contextos o tiempo. Un día cálido en invierno no es exactamente lo mismo que un día cálido en primavera. La definición exacta de cuando la temperatura va de templada a caliente es imprecisa, no podemos identificar un punto simple de templado, así que emigramos a un simple grado, la temperatura es ahora considerada caliente. Este tipo de imprecisión o difusidad asociado continuamente a los fenómenos es común en todos los campos de estudio: sociología, física, biología, finanzas, ingeniería, oceanografía, psicología, etc. (13)

En la consulta de expertos está presente la lógica difusa ya que en el cuestionario que se le da a responder a los expertos aparecen expresiones del tipo mucha relación, relación alguna relación, ninguna relación y otras las cuales tienen un grado borroso de imprecisión sin embargo de modo alguno esto no le quita científicidad a los análisis realizados ya que siempre existe un margen de error en todo tipo de investigación.

### **1.9 Descripción de las herramientas y técnicas utilizadas.**

#### **1.9.1 Software Weka.**

El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades, fue desarrollado en la Universidad de Waikato (Nueva Zelanda), implementada en Java; útil para ser aplicados sobre datos mediante las interfaces que ofrece o para insertarlos 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. Es muy útil para el desarrollo de sistemas de visión artificial (sobre todo, para tareas de extracción

de características y clasificación). Soporta varias tareas estándar de minería de datos, especialmente, pre procesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Este programa es de libre distribución y difusión. Además, ya que WEKA está programado en Java, es independiente de la arquitectura, funciona en cualquier plataforma sobre la que haya una máquina virtual java disponible. (14)

### 1.9.1.1 Herramientas y Ventajas del software Weka.

Weka tiene, entre otras, las siguientes herramientas:

- Implementaciones de numerosos clasificadores, muy útil para comprobar el comportamiento y seleccionar el más adecuado para el sistema que estemos desarrollando.
- Herramientas de clustering.
- Herramientas para selección y reducción de características, que nos permitirán, de forma sencilla, ver cuáles son más relevantes para nuestro sistema.
- Los puntos fuertes de WEKA son:
- Está disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para el pre procesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario. (15)

### 1.9.2 Statgraphics.

El programa Statgraphics es un software que está diseñado para facilitar el análisis estadístico de datos, es sencillo de aprender y utilizar gracias a su diseño intuitivo que facilita la realización de los diversos análisis. Mediante su aplicación es posible realizar un análisis descriptivo de una o varias variables, utilizando gráficos que expliquen su distribución o calculando sus medidas características. Entre sus muchas prestaciones, también figuran el cálculo de la media

aritmética, la varianza, la desviación típica y otras. En el trabajo se utiliza este analizador estadístico para dar tratamiento a las encuestas realizadas a los expertos.

La estructura de Statgraphics aporta al usuario información sobre relaciones entre datos así como gráficos a color de gran calidad, aun cuando el usuario desconozca por dónde debe comenzar a realizar el análisis. Para ello se dispone de herramientas como StatFolio y StatAdvisor, vínculos OLE, DDE y soporte ODBC y gráficos interactivos.

Con la herramienta el Asesor Estadístico (StatAdvisor) se produce de manera automática una interpretación corta y fácilmente comprensible de los resultados de los análisis estadísticos realizados, con DDE (Dynamic Data Exchange) se pueden realizar vínculos a las hojas de cálculo de Windows; mediante StatFolio se guardan y reutilizan los análisis; la Galería de Resultados (StatGallery) permite el almacenamiento de los resultados de uno o varios análisis estadísticos, generando así una presentación organizada y personalizada de los mismos, se pueden combinar textos y gráficos múltiples en varias páginas con un diseño de 32-bit que permite manejar problemas de gran magnitud y el Editor de Resultados (StatReporter) nos permitirá generar informes más personalizados que StatGallery, combinando salidas textuales y gráficas generadas por Statgraphics con nuestros propios textos. (16)

### **1.9.3 Lenguaje de programación Qt/C++.**

Qt es un producto de la empresa noruega de software Trolltech AS, esta empresa se dedica a desarrollar librerías y herramientas de desarrollo de software, además es experta en servicios de consultoría. Es un conjunto de librerías multi-plataforma para el desarrollo del esqueleto de aplicaciones GUI, escritas en código C++, además está completamente orientado a objetos y comenzó a distribuirse comercialmente en 1996 y desde entonces ha sido la base para numerosas aplicaciones incluyendo la popular interfaz gráfica para Linux llamada KDE, disponible en todas las grandes distribuciones de Linux.

En el mercado se pueden encontrar las siguientes distribuciones:

- Qt Enterprise Edition y Qt Profesional Edition, disponibles para el desarrollo de software con fines comerciales. Incluye servicio de soporte técnico y están disponibles ampliaciones.

## Capítulo 1: Fundamentación Teórica

---

- Qt Free Edition Es la versión para Unix/X11 para el desarrollo de software gratuito y de código abierto. Se puede obtener gratis sujeto a los términos de la Q Public License and the GNU General Public License. Para plataformas Windows también está disponible la versión Qt non comercial.
- Qt Educational Edition es una versión de la Qt Profesional Edition con licencia únicamente para fines educativos.
- Qt/Embedded Free Edition.

### **Características y funcionalidades.**

- QT es una librería para la creación de interfaces gráficas. Se distribuye bajo una licencia libre GPL (o QPL) que nos permite incorporar las QT en nuestras aplicaciones open-source.
- Se encuentra disponible para una gran número de plataformas: Linux, MacOS X, Solaris, HP-UX, UNIX con X11.
- Es orientado a objetos, lo que facilita el desarrollo de software. El lenguaje para el que se encuentra disponible es C++ aunque han aparecido bindings a otros lenguajes como Python o Perl.
- Es una librería que se basa en los conceptos de widgets (objetos), Señales Slots y Eventos (ejemplo: clic del ratón).
- Las señales y los slots es el mecanismo para que unos widgets se comuniquen con otros.
- Los widgets pueden contener cualquier número de hijos. El widget "top-level" puede ser cualquiera, sea ventana, botón, etc.
- Algunos atributos como el texto de etiquetas, etc. ... se modifican de modo similar al lenguaje html.

QT proporciona además otras funcionalidades:

- Librerías básicas -> Entrada/Salida, Manejo de Red, XML
- Interface con bases de datos -> Oracle, MySQL, PostgreSQL, ODBC

- Plugins, librerías dinámicas.
- Unicode, Internacionalización.

La compatibilidad de Qt, su riqueza de características, su rendimiento de C++, la disponibilidad del código fuente, su documentación de alta calidad, el apoyo técnico, y todos los elementos mencionados, hacen que los ingenieros de software disfruten de la tecnología y cumplan con sus tareas fácilmente.

### **1.9.4 Método Delphi de la consulta de expertos.**

El método Delphi consiste en la selección de un grupo de expertos a los que se les pregunta su opinión sobre cuestiones referidas a acontecimientos del futuro. Las estimaciones de los expertos se realizan en sucesivas rondas anónimas, con el objetivo de conseguir un consenso, pero con la máxima autonomía por parte de los participantes. Por lo tanto, la capacidad de predicción de Delphi se basa en la utilización sistemática de un juicio intuitivo emitido por un grupo de expertos. Este método se apoya en la interrogación a expertos seleccionados con la ayuda de cuestionarios sucesivos, a fin de poner de manifiesto convergencias de opiniones y deducir eventuales consensos. (17)

El método consta de cuatro características fundamentales:

- El anonimato.
- La iteración.
- Retroalimentación.
- Representación de las respuestas del grupo en forma estadística.

Como resultado de esas características, en el momento de aplicar este método se pueden distinguir varias fases, a continuación se definen tres fundamentales que engloban los pasos a seguir en la aplicación del mismo:

- Fase preliminar: se delimita el contexto, los objetivos, el diseño, los elementos básicos del trabajo y la selección de los expertos.
- Fase exploratoria: elaboración y aplicación de los cuestionarios según sucesivas rondas, de tal forma que con las respuestas más comunes de la primera se confecciona la siguiente.

- Fase final: análisis estadístico y presentación de la información.

El método Delphi consiste básicamente en aplicar varias rondas de preguntas, pero en la práctica generalmente no sucede así, porque con solamente una ronda o dos es suficiente para que se llegue a un consenso. (18)

Esta técnica tiene la ventaja de eliminar el efecto líder de otros métodos de expertos, pues los encuestados son anónimos entre sí, pero es muy importante la selección correcta de los expertos y definir bien el campo de investigación, con preguntas precisas, cuantificables e independientes. (19)

### Capítulo 2: Aplicación del conocimiento.

#### Introducción al capítulo.

El presente capítulo hace referencia a la fundamentación y validación de la investigación. Se hace un análisis de los diagnósticos aplicados a los estudiantes en particular a las respuestas dadas. Se confecciona la tabla de decisión del tema de investigación, aplica el software Weka en la obtención de las reglas de producción.

#### 2.1 Método Delphi de la consulta de expertos.

Como se explicó anteriormente la consulta de expertos consiste en buscar los criterios de varios especialistas de disímiles disciplinas, con el propósito de validar y disponer de una base de apoyo que respalde el diseño a utilizar. Para darle así un carácter científico al tema de investigación, ya que los niveles de conocimientos que se involucran en estos métodos son muy particulares y específicos.

Los expertos seleccionados fueron: Ingenieros y Másteres en Informática, Matemáticas y otras especialidades, todos profesores con más de 5 años de experiencia.

##### 2.1.1 Encuesta aplicada a los expertos.

###### *Cuestionario de preguntas.*

¿Cuál es su ocupación?

¿Años de experiencia?

Al inicio del curso en la UCI se realiza un diagnóstico a los estudiantes de primer año con el propósito de determinar sus principales características, por lo que se hace necesaria la obtención de un criterio relacionado con la vinculación entre las variables que se destacan.

¿Cree usted que exista relación entre las variables que se indican a continuación?

A1	Sexo	A7	Cultura Económica
A2	Inteligencia	A8	Nivel Cultural de los padres

## Capítulo 2: Aplicación del conocimiento

A3	Idioma Ingles	A9	Decisión del estudiante
A4	Lengua Materna	A10	Situación Económica
A5	Procedencia	A11	Estado de Salud
A6	Matemática	A12	Estilo Decisinal
		A13	Perfil de Inteligencia.

¿Si usted cree que existe relación entre las variables que se indican a continuación? Marque con una cruz (X) el grado de relación que usted cree que tienen las variables. Por ejemplo A1-A2 significa que el sexo depende de la inteligencia y A2-A1 significa lo contrario.

Variables		Mucha relación		Relación		Alguna relación		Poca relación		Ninguna relación	
A1 – A2	A2– A1										
A1- A3	A3- A1										
A1– A4	A4- A1										
A6– A2	A2- A6										
A7– A5	A5- A7										
A4– A5	A5- A4										
A6– A5	A5- A6										
A8– A2	A2- A8										
A8– A4	A4- A8										
A8– A6	A6- A8										

## Capítulo 2: Aplicación del conocimiento

---

A9- A5	A5- A9											
A9- A2	A2- A9											
A9- A7	A7- A9											
A9- A3	A3- A9											
A10- A2	A2- A10											
A11- A13	A13 -A11											
A11- A5	A5- A11											
A11- A1	A1- A11											
A12- A5	A5- A12											
A12- A10	A10- A12											
A1- A12	A12- A1											
A12- A7	A7- A12											

### 2.1.2 Resultados obtenidos.

Una encuesta puede interpretarse de varias formas y puede provocar que el experto no siempre comprenda lo que se le pregunta, con el propósito de evitar este problema el cual ocurre con frecuencia y para poder obtener una información altamente confiable en el momento de aplicar la encuesta, fueron hechas las siguientes aclaraciones:

El criterio se toma de forma bidireccional y existe una casilla para cada respuesta.

En caso de existir dudas en cuanto a qué criterio debería marcarse en alguno de los pares propuestos, existe la posibilidad de no elegir ninguno, el experto puede dejarlo en blanco.

## Capítulo 2: Aplicación del conocimiento

Luego de aplicada la encuesta a los expertos seleccionados, se procedió a la elaboración de una tabla con el objetivo de resumir los resultados obtenidos. Para la confección de la misma se realizó un conteo por cada par de variables propuestas.

La cantidad de expertos que optaron por cada criterio, no siempre es numéricamente igual a la cantidad de expertos encuestados, esto se debe a la segunda aclaración antes mencionada. Luego de realizar el conteo, no todos los expertos tuvieron el mismo criterio respecto al tipo de relación que existe entre cada par de variables. A continuación la tabla 3 muestra los resultados obtenidos.

Variables		Mucha relación		Relación		Alguna Relación		Poca relación		Ninguna relación	
A1 – A2	A2– A1									X	X
A1- A3	A3- A1									X	X
A1– A4	A4- A1									X	X
A6– A2	A2- A6	X	X								
A7– A5	A5- A7	X	X								
A4– A5	A5- A4	X	X								
A6– A5	A5- A6									X	X
A8– A2	A2- A8			X	X						
A8– A4	A4- A8	X	X								
A8– A6	A6- A8			X	X						
A9– A5	A5- A9							X			X
A9– A2	A2- A9			X	X						

## Capítulo 2: Aplicación del conocimiento

A9- A7	A7- A9					X	X					
A9- A3	A3- A9									X	X	
A10- A2	A2- A10						X			X		
A11- A13	A13-A11									X	X	
A11- A5	A5- A11									X	X	
A11- A1	A1- A11									X	X	
A12- A5	A5- A12					X	X					
A12- A10	A10-A12				X	X						
A1- A12	A12- A1					X	X					
A12- A7	A7- A12	X	X									

**Tabla 3: Resultados obtenidos en la encuesta aplicada a los expertos.**

A partir del análisis de la tabla anterior, fue obtenida la matriz booleana representada en la tabla 4. Para completar los valores de cada posición, se tomó la decisión de poner 1 en caso de existir mucha relación o relación, en otro caso se puso 0.

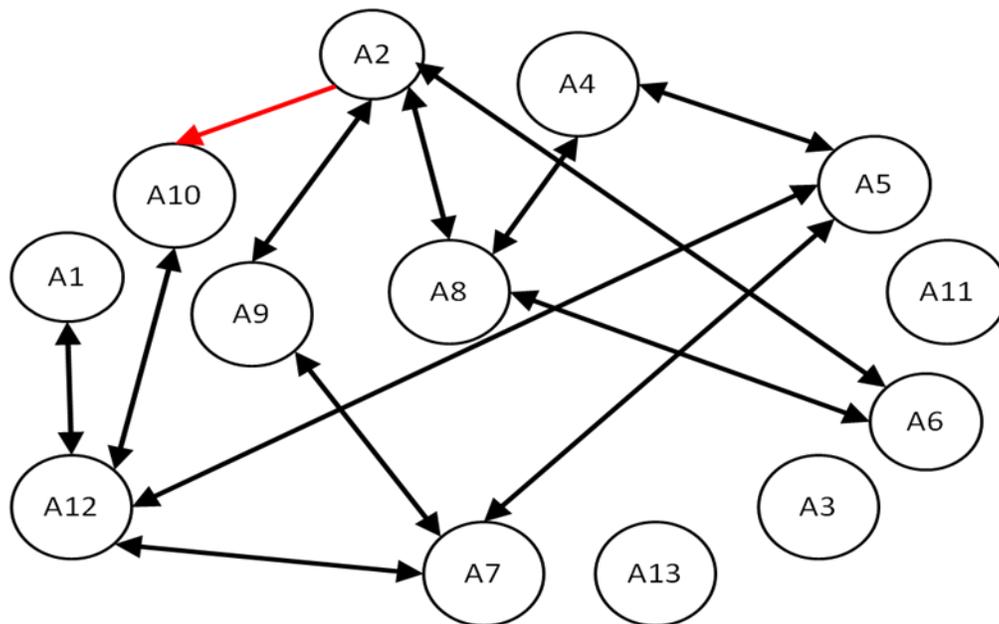
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
A1	0	0	0	0	0	0	0	0	0	0	0	1	0
A2	0	0	0	0	0	1	0	1	1	1	0	0	0
A3	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	0	1	0	0	1	0	0	0	0	0
A5	0	0	0	1	0	0	1	0	0	0	0	1	0

## Capítulo 2: Aplicación del conocimiento

A6	0	1	0	0	0	0	0	1	0	0	0	0	0
A7	0	0	0	0	1	0	0	0	1	0	0	1	0
A8	0	1	0	1	0	1	0	0	0	0	0	0	0
A9	0	1	0	0	0	0	1	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	0	0	1	0
A11	0	0	0	0	0	0	0	0	0	0	0	0	0
A12	1	0	0	0	1	0	1	0	0	1	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	0

**Tabla 4: Matriz Booleana.**

Para conocer cuáles son las variables que más relaciones tienen dentro del entorno, es conformado el grafo que se muestra en la figura 4 donde se observan las variables y sus relaciones correspondientes, a partir de los resultados obtenidos de la matriz booleana. En este caso se definen como nodos o vértices los elementos: Sexo (A1), Inteligencia (A2), Idioma Ingles (A3), Lengua Materna(A4), Procedencia (A5), Matemática (A6), Cultura Económica (A7), Nivel Cultural de los padres(A8), Decisión del estudiante(A9), Situación Económica(A10), Estado de Salud (A11), Estilo Decisional (A12), Perfil de Inteligencia( A13). Las saetas de doble sentido indican la existencia de una relación bidireccional entre el par de variables, donde la correspondencia entre ellas es mucha relación, relación y alguna relación y la de un solo sentido que la variable origen solo tiene relación con la destino.



**Figura 4: El grafo muestra la relación entre las variables.**

La tabla 5, muestra los poderes de cada vértice. En la misma se determinan los poderes de tipo A y B de las variables involucradas y la suma de estos representa el peso total de cada variable. El de tipo A se determina por la suma de las columnas y el de tipo B por la suma de las filas. Determinado el poder de cada una, se puede definir cuáles son las variables más importantes, es oportuno destacar que el poder de una variable es directamente proporcional al grado de importancia de la misma.

Variables	Poder A	Poder B	Poder A + Poder B
A1	1	1	2
A2	3	4	7
A3	0	0	0
A4	2	2	4
A5	3	3	6
A6	2	2	4

A7	3	3	6
A8	3	3	6
A9	2	2	4
A10	1	1	2
A11	0	0	0
A12	4	4	8
A13	0	0	0

**Tabla 5: Tabla de Poderes.**

Las variables con más peso, según el criterio de los expertos, son A12 y A2 correspondientes a: Estilo Decisional e Inteligencia.

Todos los valores obtenidos tienen como punto de partida las preguntas realizadas a los expertos donde las respuestas mucha relación, relación, alguna relación, poca relación y ninguna relación presentan cierto grado de incertidumbre y están vinculadas a la lógica difusa ya que no se puede determinar con exactitud la diferencia entre los parámetros anteriormente mencionados.

### **2.1.3 Análisis estadístico de la información.**

Para validar los resultados obtenidos de la encuesta, es necesario realizar un análisis estadístico (Ver Tabla 6), para esto se utilizó el asistente estadístico Statgraphics. Dicho análisis mostrará el nivel de concordancia existente entre los expertos según las repuestas dada y si es necesario aplicar una segunda ronda de encuesta.

Se determina la media aritmética de los criterios de los expertos para cada elemento, obteniendo el coeficiente de variación por medio del cálculo de la Desviación Típica (S).

Si  $S > 1$ , se rechaza el valor promedio calculado y se realiza una nueva ronda de preguntas.

Si  $S < 1$ , se acepta el criterio de los expertos.

## Capítulo 2: Aplicación del conocimiento

Si S=1, significa acuerdo total entre los expertos.

Variables	Cantidad	Promedio	Varianza	Desviación Típica(S)	Mínimo	Máximo	Rango
A1 – A2	15	1.9	1.0	1.0	1	4	3
A1- A3	14	1.4	0.2	0.5	1	2	1
A1– A4	14	1.7	0.9	0.9	1	4	3
A6– A2	14	4.3	0.7	0.8	3	5	2
A7– A5	15	4.3	0.8	0.9	3	5	2
A4– A5	15	4.2	0.9	0.9	3	5	2
A6– A5	14	2.3	1.3	1.1	1	4	3
A8– A2	15	3.1	1.2	1.1	1	5	4
A8– A4	14	4.3	0.7	0.8	3	5	2
A8– A6	14	3.6	0.9	0.9	1	5	4
A9– A5	15	2.7	1.1	1.0	1	5	4
A9– A2	15	3.5	0.7	0.8	2	5	3
A9– A7	14	3.2	1.0	1.0	1	4	3
A9- A3	15	1.7	1.5	1.0	1	5	4
A10- A2	14	1.9	1.3	1.1	1	5	4
A11- A13	14	1.8	1.4	1.1	1	5	4
A11- A5	14	1.8	1.1	1.0	1	4	3

## Capítulo 2: Aplicación del conocimiento

A11- A1	15	1.9	1.3	1.1	1	4	3
A12- A5	15	3.5	0.7	0.8	2	5	3
A12- A10	14	3.9	0.7	0.8	3	5	2
A1- A12	15	2.7	0.9	1.0	1	4	3
A12- A7	15	4.2	0.8	0.9	3	5	2
A2- A1	13	1.6	0.8	0.9	1	3	2
A3- A1	13	1.4	0.3	0.5	1	2	1
A4- A1	13	1.8	1.2	1.0	1	4	3
A2- A6	13	3.7	1.2	1.1	2	5	3
A5- A7	12	4.2	1.2	1.0	2	5	3
A5- A4	13	4.1	0.9	0.9	3	5	2
A5- A6	13	1.8	1.1	1.0	1	4	3
A2- A8	13	2.9	1.1	1.0	1	4	3
A4- A8	13	4.2	0.8	0.9	3	5	2
A6- A8	14	3.2	1.3	1.1	1	4	3
A5- A9	12	1.8	0.7	0.8	1	3	2
A2- A9	15	2.6	1.0	1.0	1	5	4
A7- A9	13	2.5	1.3	1.1	1	4	3
A3- A9	14	1.9	1.3	1.1	1	4	3

A2- A10	14	2.4	1.2	1.0	1	4	3
A13 -A11	14	1.9	1.0	1.0	1	4	3
A5- A11	15	2.1	1.2	1.0	1	4	3
A1- A11	14	1.5	0.9	0.9	1	4	3
A5- A12	10	3.1	1.0	1.0	1	4	3
A10- A12	11	3.8	0.6	0.7	3	5	2
A12- A1	15	2.4	1.1	1.0	1	4	3
A7- A12	13	3.5	1.1	1.0	2	5	3

**Tabla 6. Resultados obtenidos del análisis estadístico.**

La cantidad de respuestas dadas por cada relación oscila entre 10 y 15, esto es debido a que algunos expertos decidieron no dar un criterio a algunas relaciones, esto no influye en la veracidad de la encuesta. Luego de realizar el tratamiento estadístico y de comparar la Desviación Típica (S) con las propuestas antes expuestas se acepta el criterio dado por los expertos, los casos que no satisfacen lo dicho anteriormente también se aceptan debido a que la diferencia por la que sobrepasan este valor no es significativa.

### 2.1.4 Gráficos obtenidos.

Las siguientes gráficas representan el porcentaje de concordancia de los expertos al responder la relación que existía entre los pares de variable tomadas como ejemplo. El eje vertical muestra el porcentaje de concordancia y el eje horizontal muestra los criterios que se tuvieron en cuenta: Ninguna Relación (NR), Poca Relación (PR), Alguna Relación (AR), Relación (R), Mucha Relación (MR).

## Capítulo 2: Aplicación del conocimiento

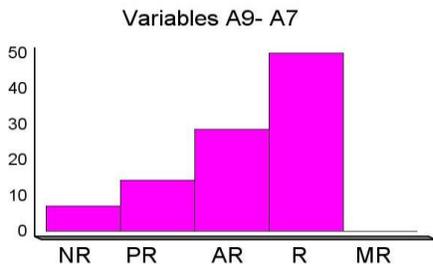


Figura 5: Decisión del Estudiante-

Cultura Económica

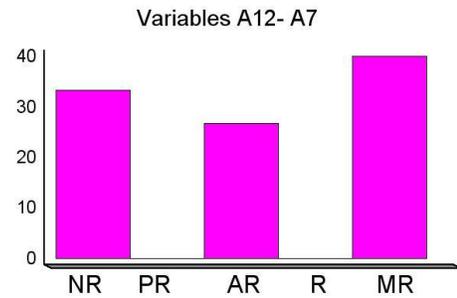


Figura 6: Estilo Decisional-

Cultura Económica

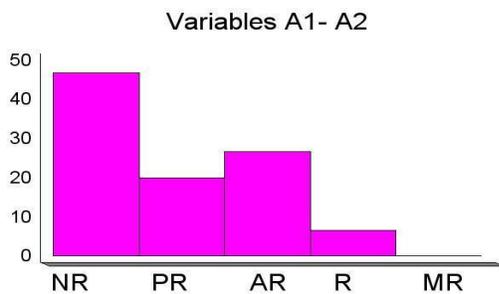


Figura 7: Sexo- Inteligencia.

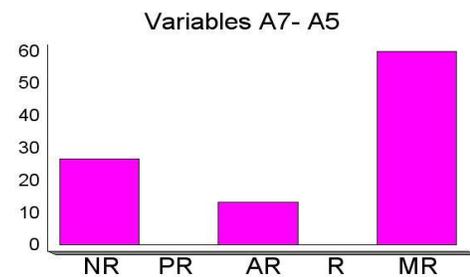


Figura 8: Cultura Económica-

Procedencia.

El gráfico de barra que se muestra a continuación es una representación del total de respuestas obtenidas y la gráfica de pasteles el porcentaje que representa de la misma.

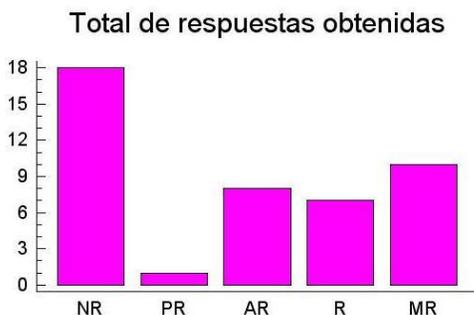


Figura 9: Total de respuestas de los expertos

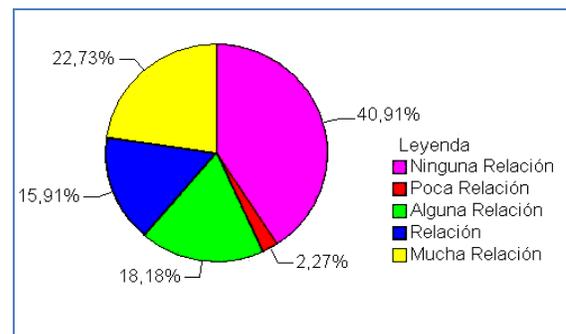


Figura 10: Representación del por ciento de la concordancia de los expertos.

### 2.2 Proceso de descubrimiento del conocimiento. Análisis de los estudiantes encuestados.

En la actualidad la existencia de voluminosas Bases de Hechos conteniendo grandes cantidades de datos es una realidad en muchas organizaciones y esto trae consigo que frecuentemente se tomen decisiones teniendo en cuenta la intuición y experiencia del decisor, más que considerando la rica información almacenada.

Durante el proceso de investigación se confeccionó una Base de Hecho perteneciente a los resultados obtenidos en los 7 diagnósticos aplicados a los estudiantes de primer año. Para la confección de la misma se realizó un análisis minucioso de cuántos estudiantes respondieron los 7 diagnósticos. Esta información se almacenó en un matriz (Ver Tabla 7) conformada por filas y columnas donde las filas representan los 115 estudiantes que respondieron y las columnas los nombres de los 7 diagnósticos. El total de estudiantes representa un 70% de la matrícula de primer año que es una muestra representativa.

#	Nombres y Apellidos	Estilo de Aprendizaje	Estrategia Aprendizaje	Perfi de Inteligencia	Cultura General e Ideopolítico	Idioma Inglés	Lengua Materna	Matemática y Algoritmización	Grupo
1	A1	x	x	x	x	x	x	x	5101
2	A2	x	x	x	x	x	x	x	.....
3	A3	x	x	x	x	x	x	x	.....
...	.....	x	x	x	x	x	x	x	.....
...	.....	x	x	x	x	x	x	x	.....
...	.....	x	x	x	x	x	x	x	.....
...	.....	x	x	x	x	x	x	x	.....
...	.....	x	x	x	x	x	x	x	.....
115	A115	x	x	x	x	x	x	x	5106

Tabla 7. Relación de estudiantes que respondieron los 7 diagnósticos.

#### 2.2.1 Análisis de los Diagnósticos.

Para seleccionar los atributos más importantes y el valor del domino que pueden tomar, fue necesario hacer un análisis de cada atributo existente en las encuestas. Para almacenar la información se confeccionó una tabla para cada encuesta, las mismas en las filas almacenan

## Capítulo 2: Aplicación del conocimiento

los 115 estudiantes que respondieron el diagnóstico y por las columnas la cantidad de puntos que obtuvo cada estudiante en las distintas secciones, representada por RS# (Respuesta de la Sección 1), y 20-30(rango de las preguntas asociadas), además del nombre de la encuesta, la cantidad de puntos máximos a obtener en las secciones y el promedio.

A continuación se muestra la información obtenida para el Diagnóstico Perfil de Inteligencia.

Perfil de Inteligencia									
Puntos máximos por sección	50	50	50	50	50	50	50	50	
Respuestas Secciones	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS8	Promedio
Nombre Estudiante	(1-10)	(11-20)	(21-30)	(31-40)	(41-50)	(51-60)	(61-70)	(71-80)	
Estudiante 1	20	17	20	18	9	19	9	8	15
Estudiante 2	20	35	30	31	20	26	38	24	28
Estudiante 3	44	49	44	45	33	41	45	48	44
Estudiante 4	34	32	41	44	40	28	39	34	37
Estudiante 5	26	38	44	35	31	32	41	31	35
Estudiante 6	32	39	40	41	33	36	44	27	37
Estudiante 7	38	43	47	42	40	48	44	46	44
Estudiante 8	32	34	36	45	26	36	37	25	34
Estudiante 9	35	45	50	47	39	47	40	43	43
Estudiante 10	31	41	37	37	38	38	47	42	39

## Capítulo 2: Aplicación del conocimiento

Estudiante 11	35	44	41	49	23	41	42	40	39
Estudiante 12	34	43	31	45	30	37	50	35	38
Estudiante 13	38	38	33	42	39	42	39	30	38
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Estudiante 115	44	33	31	37	45	35	42	46	39

**Tabla 8. Resultados obtenidos en el diagnóstico Perfil de Inteligencia.**

En los diagnósticos Estrategia de Aprendizaje, Lengua Materna, Cultura General e Ideopolítico, Idioma Inglés, Estilo de Aprendizaje y Matemática y Algoritmización los valores máximos de las secciones son diferentes. Con el propósito de uniformar los resultados como aparecen en la primera encuesta cada celda se multiplicó por un factor  $K=50/PM$  siendo PM la puntuación máxima. En la última columna se refleja el promedio de cada una de las notas obtenidas. A continuación se muestran las tablas con los resultados obtenidos.

Diagnóstico Estrategia de Aprendizaje									
Puntos máximos por sección	10	20	10	10	10	10	10	10	
Respuestas Secciones	RS1	RS2	RS3	RS4	RS5	RS6	RS7	RS8	Promedio
Nombre Estudiante	(1-5)	(6-15)	(16-20)	(21-25)	(26-30)	(31-35)	(36-40)	(41-45)	
Estudiante 1	7	11	7	5	5	1	7	6	23
Estudiante 2	7	5	7	6	8	5	8	4	28
Estudiante 3	9	18	10	9	3	8	10	10	41

## Capítulo 2: Aplicación del conocimiento

Estudiante 4	8	19	9	10	6	6	10	10	40
Estudiante 5	9	14	9	9	6	6	8	8	38
Estudiante 6	10	13	9	8	5	7	6	8	38
Estudiante 7	6	14	6	8	5	6	10	7	32
Estudiante 8	9	15	9	6	7	5	7	9	36
Estudiante 9	8	11	9	10	4	6	6	5	33
Estudiante 10	8	17	10	5	7	7	10	9	38
Estudiante 11	4	17	6	10	9	3	4	9	33
Estudiante 12	7	18	5	5	10	7	7	10	38
Estudiante 13	3	14	9	7	7	6	7	8	33
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Estudiante 115	9	17	9	10	5	8	0	9	42

**Tabla 9. Resultados obtenidos en el diagnóstico Estrategia de Aprendizaje.**

Lengua Materna				
Puntos máximos por sección	9	12	9	
Respuestas Secciones	RS1	RS2	RS3	
Nombre Estudiante	(1-9)	(10-21)	(22-30)	Promedio
Estudiante 1	9	10	8	45

## Capítulo 2: Aplicación del conocimiento

Estudiante 2	7	8	8	45
Estudiante 3	3	6	5	28
Estudiante 4	6	10	6	34
Estudiante 5	6	9	8	45
Estudiante 6	8	11	7	39
Estudiante 7	6	9	7	39
Estudiante 8	5	8	8	45
Estudiante 9	7	9	5	28
Estudiante 10	6	4	5	28
Estudiante 11	7	8	8	45
Estudiante 12	8	11	8	45
Estudiante 13	7	9	7	39
.....	.....	.....	.....	.....
Estudiante 115	6	0	0	0

**Tabla 10. Resultados obtenidos en el diagnóstico Lengua Materna.**

General e Ideo-político				
Puntos máximos por sección	20	12	28	
Respuestas Secciones				
Nombre Estudiante	RS1(1-10)	RS2(1-6)	RS3(1-14)	Promedio

## Capítulo 2: Aplicación del conocimiento

Estudiante 1	12	8	28	31
Estudiante 2	14	8	20	30
Estudiante 3	12	6	22	26
Estudiante 4	18	12	28	41
Estudiante 5	16	12	24	38
Estudiante 6	14	12	24	36
Estudiante 7	10	8	22	27
Estudiante 8	12	10	28	33
Estudiante 9	16	8	22	32
Estudiante 10	14	10	26	34
Estudiante 11	14	8	26	32
Estudiante 12	16	12	26	39
Estudiante 13	12	10	22	31
.....	.....	.....	.....	.....
Estudiante 115	18	8	26	35

**Tabla 11. Resultados obtenidos en el diagnóstico General e Ideo-político.**

Idioma Inglés						
Puntos máximos por sección	12	12	12	45	19	
<del>Respuestas</del> Secciones	RS1	RS2	RS3	RS4	RS5	Promedio

## Capítulo 2: Aplicación del conocimiento

Nombre Estudiante	(1-6)	(7-12)	(13-18)	(19-27)	(28-30)	
Estudiante 1	6	8	4	35	14	19
Estudiante 2	8	6	4	35	14	19
Estudiante 3	12	12	8	40	14	34
Estudiante 4	12	12	4	35	7	28
Estudiante 5	4	6	4	20	7	15
Estudiante 6	10	6	0	0	0	15
Estudiante 7	4	4	8	10	1	18
Estudiante 8	2	4	2	15	0	9
Estudiante 9	10	10	4	20	1	24
Estudiante 10	8	8	0	20	1	15
Estudiante 11	4	2	6	20	9	14
Estudiante 12	10	10	6	40	17	27
Estudiante 13	12	10	4	35	16	26
.....	.....	.....	.....	.....	.....	.....
Estudiante 115	2	4	8	20	9	17

**Tabla 12. Resultados obtenidos en el diagnóstico Idioma Inglés.**

Diagnóstico Estilo de Aprendizaje

## Capítulo 2: Aplicación del conocimiento

Puntos máximos por sección	10	10	10	10	10	10	10	10	
Respuestas Secciones	RS1 (1-5)	RS2 (6-10)	RS3 (11-15)	RS4 (16-20)	RS5 (21-25)	RS6 (26-30)	RS7 (31-35)	RS8 (36-40)	Promedio
Nombre Estudiante									
Estudiante 1	7	6	6	6	8	4	7	4	30
Estudiante 2	7	5	7	6	8	5	8	4	31
Estudiante 3	7	3	6	10	10	4	9	7	35
Estudiante 4	8	4	4	7	9	7	3	7	31
Estudiante 5	7	3	5	7	9	7	6	4	30
Estudiante 6	5	4	5	5	8	5	6	6	28
Estudiante 7	9	3	9	10	10	7	7	7	39
Estudiante 8	7	5	5	7	10	2	4	4	28
Estudiante 9	7	4	7	9	10	4	5	4	31
Estudiante 10	8	7	7	7	8	5	10	2	34
Estudiante 11	9	8	6	8	10	5	4	5	34
Estudiante 12	8	7	8	3	7	5	5	7	31
Estudiante 13	9	3	6	7	10	4	5	7	32
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Estudiante 115	8	3	5	9	8	9	7	2	32

**Tabla 13. Resultados obtenidos en el diagnóstico Estilo de Aprendizaje.**

Diagnóstico de Matemática y Algoritmización						
Puntos máximos por sección	40	32	56	12	20	
Respuestas Secciones						
Nombre Estudiante	RS1(1-12)	RS2(13-20)	RS3(23-40)	RS4(41-43)	RS5(44-49)	Promedio
Estudiante 1	15	0	0	2	10	11
Estudiante 2	2	2	0	2	5	6
Estudiante 3	15	0	0	0	0	4
Estudiante 4	20	17	1	7	5	19
Estudiante 5	20	0	0	17	2	20
Estudiante 6	25	0	0	0	0	7
Estudiante 7	15	0	16	2	5	11
Estudiante 8	30	2	0	0	5	11
Estudiante 9	10	12	0	12	10	21
Estudiante 10	10	0	0	0	0	3
Estudiante 11	10	0	0	0	0	3
Estudiante 12	0	0	2	0	0	1

## Capítulo 2: Aplicación del conocimiento

Estudiante 13	10	22	16	2	15	22
.....	.....	.....	.....	.....	.....	.....
Estudiante 115	20	2	16	2	0	10

**Tabla 14: Resultados obtenidos en el diagnóstico de Matemática y Algoritmización.**

Durante el procesamiento de las encuestas se observó que el CICE para caracterizar a sus estudiantes no tiene en cuenta la Escolaridad de los Padres (EP), la Procedencia del Estudiante (PE), el Lugar donde viven (Zona) y su Situación Económica (SE), factores importantes que influyen en el proceso de aprendizaje de los estudiantes; por la importancia e información que estos aportan a la hora de de caracterizar a los educandos se adicionan a la investigación con el objetivo de que el profesor obtenga una característica mas completa del estudiante y utilice nuevos métodos que ayuden en el proceso de aprendizaje de los mismo. Para cada atributo se le asigna un rango de categoría y a estos se les asocia un peso el cual presenta cierto grado de incertidumbre ya que no se puede determinar con exactitud la diferencia entre los parámetros (Ver Tabla 15).

Situación Económica		Escolaridad de los Padres		Zona		Procedencia del Estudiante	
Muy Bien (MB)	8	Universitario(U)	9	Urbana	8	Escuela Vocacional (IPVCE)	8
Bien(B)	7	Técnico Medio(T)	7	Rural	6	Escuela Militar(EMCC )	7
Regular(R)	6	Bachiller(B)	6			Preuniversitario(IPUEC)	6
Mal(M)	4	Obrero Calificado (O)	5			Técnico Medio en Informática(TMI)	6
		Secundaria Básica(S)	4			Escuela de Deporte(Deporte)	5

**Tabla 15. Categoría y peso de cada atributo.**

Una muestra de cómo quedó organizada la información que contenían los 4 atributos se muestran en la tabla 16, donde por ejemplo: U-U significa que tanto el padre como la madre son graduados Universitarios.

## Capítulo 2: Aplicación del conocimiento

Nombres y Apellidos	Situación Económica	Escolaridad de los Padres	Zona (Rural o Urbana)	Procedencia del Estudiante
Estudiante 1	R	U-U	Rural	T MI
Estudiante 2	R	U-O	Urbana	TMI
Estudiante 3	B	T-T	Urbana	IPUEC
Estudiante 4	B	U-T	Rural	TMI
Estudiante 5	B	B-U	Urbana	IPUEC
Estudiante 6	B	T-T	Urbana	IPU
Estudiante 7	B	T-U	Urbana	IPU
Estudiante 8	R	U-U	Urbana	IPVCE
Estudiante 9	R	B-S	Rural	DEPORTE
Estudiante 10	B	U-U	Urbana	IPUEC
Estudiante 11	B	U-U	Urbana	IPUEC
Estudiante 12	R	U-U	Urbana	TM
Estudiante 13	B	U-U	Urbana	IPVCE
.....	.....	.....	.....	.....
Estudiante 115	B	U-U	Urbana	IPVCE

**Tabla 16. Resultados obtenidos de los atributos adicionados a la investigación.**

Para la obtención de esta tabla se sustituyeron los valores de cada instancia y se les aplicó la función explicada con anterioridad siendo  $PM=10$ . Los resultados obtenidos luego de realizar este proceso se muestran en la siguiente tabla.

## Capítulo 2: Aplicación del conocimiento

Nombres y Apellidos	Situación Económica	Escolaridad de los Padres	Zona (Rural o Urbana)	Procedencia del Estudiante
Estudiante 1	30	45	30	30
Estudiante 2	30	30	40	30
Estudiante 3	35	35	40	30
Estudiante 4	35	40	30	30
Estudiante 5	35	35	40	30
Estudiante 6	35	30	40	30
Estudiante 7	35	40	40	30
Estudiante 8	30	45	40	35
Estudiante 9	30	25	30	30
Estudiante 10	35	45	40	30
Estudiante 11	35	45	40	30
Estudiante 12	30	45	40	30
Estudiante 13	35	45	40	35
.....	.....	.....	.....	.....
Estudiante 115	35	45	40	35

**Tabla 17. Identificadores de los atributos.**

### 2.2.2 Base de Hechos.

Para la elaboración de la base de hechos se utilizaron los valores de los promedio obtenidos anteriormente en cada una de las encuestas, la misma consta de 12 atributos y 115 instancias (muestra tomada para la investigación), que serán utilizadas para la obtención del conocimiento. La tabla 18 a continuación muestra los atributos con su identificador y los posibles valores que pueden tomar, así como, el orden en el que aparecen en la Base de Hechos.

---

ID	Atributo	Valor de Domino
B1	Estilo Decisional	0-19,20-29,30-34,35-39,40-44,45-50
B2	Estrategia de Aprendizaje	0-19,20-29,30-34,35-39,40-44,45-50
B3	Perfil de Inteligencia;	0-19,20-29,30-34,35-39,40-44,45-50
B4	Lengua Materna	0-19,20-29,30-34,35-39,40-44,45-50
B5	Idioma Inglés	0-19,20-29,30-34,35-39,40-44,45-50
B6	Matemática y Algoritmización	0-19,20-29,30-34,35-39,40-44,45-50
B7	Cultura Económica	0-19,20-29,30-34,35-39,40-44,45-50
B8	Situación Económica	Muy Bien, Bien, Regular, Mal

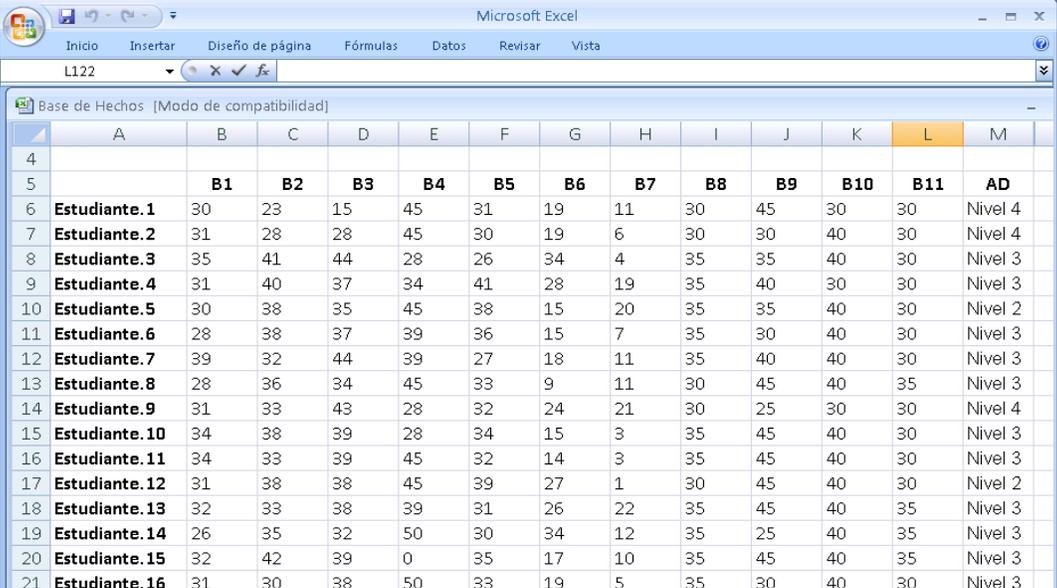
---

## Capítulo 2: Aplicación del conocimiento

B9	Escolaridad de los Padres	Universitario, Técnico Medio, Bachiller, Obrero Calificado, Secundaria Básica
B10	Zona	Urbana Rural
B11	Procedencia del Estudiante	IPVCE, IPUEC, TMI, EMCC, Deporte
AD	Atributo de Decisión	Nivel1,Nivel2,Nivel3,Nivel4,Nivel5

**Tabla 18. Clasificación de los atributos de la Base de Hechos**

La Base de Hechos fue creada en una plantilla Excel donde se almacenan los valores de las instancias, una muestra de ello en la tabla a continuación.



	A	B	C	D	E	F	G	H	I	J	K	L	M
4													
5		<b>B1</b>	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>B8</b>	<b>B9</b>	<b>B10</b>	<b>B11</b>	<b>AD</b>
6	<b>Estudiante.1</b>	30	23	15	45	31	19	11	30	45	30	30	Nivel 4
7	<b>Estudiante.2</b>	31	28	28	45	30	19	6	30	30	40	30	Nivel 4
8	<b>Estudiante.3</b>	35	41	44	28	26	34	4	35	35	40	30	Nivel 3
9	<b>Estudiante.4</b>	31	40	37	34	41	28	19	35	40	30	30	Nivel 3
10	<b>Estudiante.5</b>	30	38	35	45	38	15	20	35	35	40	30	Nivel 2
11	<b>Estudiante.6</b>	28	38	37	39	36	15	7	35	30	40	30	Nivel 3
12	<b>Estudiante.7</b>	39	32	44	39	27	18	11	35	40	40	30	Nivel 3
13	<b>Estudiante.8</b>	28	36	34	45	33	9	11	30	45	40	35	Nivel 3
14	<b>Estudiante.9</b>	31	33	43	28	32	24	21	30	25	30	30	Nivel 4
15	<b>Estudiante.10</b>	34	38	39	28	34	15	3	35	45	40	30	Nivel 3
16	<b>Estudiante.11</b>	34	33	39	45	32	14	3	35	45	40	30	Nivel 3
17	<b>Estudiante.12</b>	31	38	38	45	39	27	1	30	45	40	30	Nivel 2
18	<b>Estudiante.13</b>	32	33	38	39	31	26	22	35	45	40	35	Nivel 3
19	<b>Estudiante.14</b>	26	35	32	50	30	34	12	35	25	40	35	Nivel 3
20	<b>Estudiante.15</b>	32	42	39	0	35	17	10	35	45	40	35	Nivel 3
21	<b>Estudiante.16</b>	31	30	38	50	33	19	5	35	30	40	30	Nivel 3

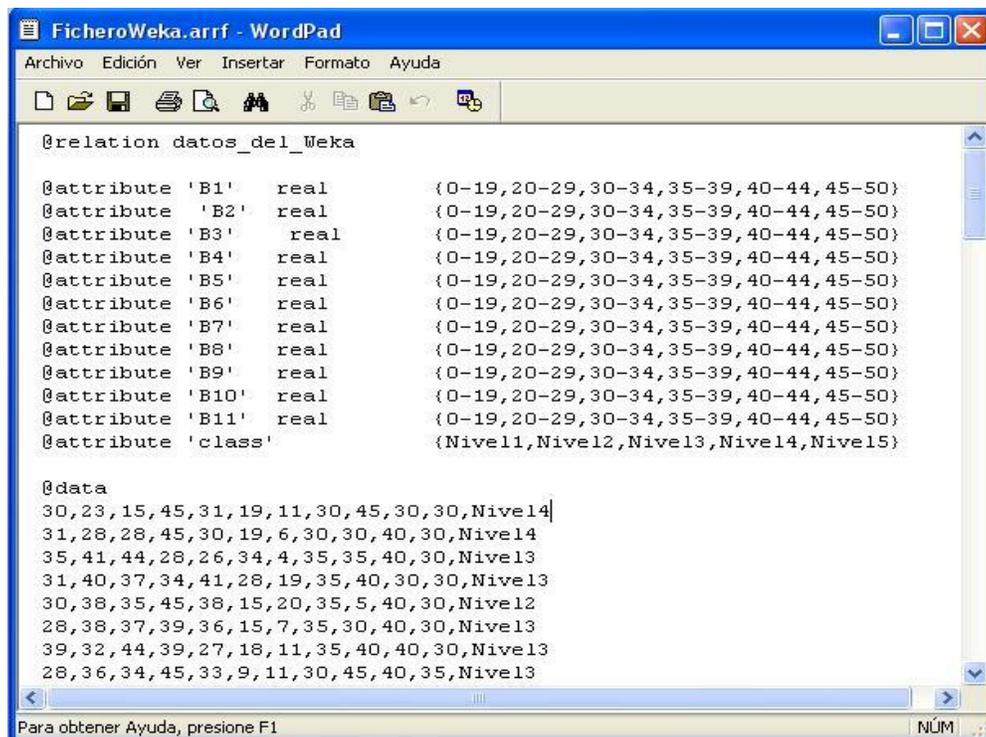
**Tabla 18. Datos de la Base de Hecho.**

### **2.3 Procesamiento de datos.**

El procesamiento de datos es el paso principal cuando se necesita obtener información de un amplio conjunto de datos. En este caso el proceso consta de dos fases importantes: la preparación de los datos a procesar y el procesamiento. Una vez que la Base de Hechos fue creada y revisada se procedió al análisis de los datos utilizando el software WEKA.

#### **2.3.1 Preparación de los datos en el WEKA.**

La primera acción que se realizó fue convertir la Base de Hechos en un fichero extensión .arff (Attribute-Relation File Format) que es la soportada por el software a utilizar. Para esto se salvaron todos los datos de la Base de Hechos a un archivo WordPad con nombres Fichero WEKA y se guardó con la extensión especificada. Como parte del proceso de preparación de los datos se eliminaron inconsistencias de forma manual, tales como: errores al pasar valores de los atributos y la eliminación de instancias que se habían duplicado. Este paso se considera de vital importancia debido a que su correcta aplicación impacta de manera directa en el conocimiento a descubrir. En el fichero a los atributos se les especifica los valores que pueden tomar. A continuación se muestra como quedó el fichero .arff.



```
@relation datos_del_Weka
@attribute 'B1' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B2' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B3' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B4' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B5' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B6' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B7' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B8' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B9' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B10' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'B11' real {0-19,20-29,30-34,35-39,40-44,45-50}
@attribute 'class' {Nivel11,Nivel12,Nivel13,Nivel14,Nivel15}
@data
30,23,15,45,31,19,11,30,45,30,30,Nivel14
31,28,28,45,30,19,6,30,30,40,30,Nivel14
35,41,44,28,26,34,4,35,35,40,30,Nivel13
31,40,37,34,41,28,19,35,40,30,30,Nivel13
30,38,35,45,38,15,20,35,5,40,30,Nivel12
28,38,37,39,36,15,7,35,30,40,30,Nivel13
39,32,44,39,27,18,11,35,40,40,30,Nivel13
28,36,34,45,33,9,11,30,45,40,35,Nivel13
```

Figura 11: Archivos .arff.

### 2.3.2 Procesamiento de la Base de Hechos y resultados obtenidos.

Una vez creado el archivo .arff se procede al procesamiento de los datos con el software WEKA.

El primer paso para el procesamiento de esta Base de Hechos fue cargar el fichero en el software, donde se visualizan las características de los atributos. A esta primera acción se le llama pre procesamiento. El próximo paso fue realizar el procesamiento de los datos. Para esto se seleccionó la pestaña Classify y dentro de ella se eligió el clasificador Trees que es el correspondiente a los atributos de decisión y se eligió el algoritmo J48graft para el procesamiento (Ver Anexo 1). El resultado que se muestra en la figura, representa un conjunto de primicias escritas en forma de escalón, las cuales simbolizan el atributos de decisión y por tanto el conocimiento que se desea obtener. Luego de realizar todo este proceso se aplicó el procesamiento obteniéndose el siguiente resultado:

```

== Classifier model (full training set) ==
J48graft pruned tree

B2 <= 31
| B4 <= 45: Nivel4 (26.0/5.0)
| B4 > 45: Nivel3 (5.0/1.0)
B2 > 31
| B8 <= 35
| | E10 <= 30
| | | B5 <= 34
| | | | B9 <= 30: Nivel4 (6.0)
| | | | B9 > 30
| | | | | B4 <= 34: Nivel4 (2.0)
| | | | | B4 > 34: Nivel3 (3.0)
| | | | B5 > 34: Nivel3 (7.0)
| | | B10 > 30
| | | | B5 <= 36
| | | | | B11 <= 30
| | | | | | B9 <= 25: Nivel4 (3.0)
| | | | | | B9 > 25
| | | | | | | B8 <= 30
| | | | | | | | B2 <= 33: Nivel4 (2.0)
| | | | | | | | B2 > 33: Nivel3 (4.0/1.0)
| | | | | | | B8 > 30
| | | | | | | | B4 <= 34
| | | | | | | | | B9 <= 35
| | | | | | | | | | B6 <= 26: Nivel4 (2.0)
| | | | | | | | | | B6 > 26: Nivel3 (2.0)
| | | | | | | | | B9 > 35: Nivel3 (3.0)
| | | | | | | | B4 > 34
| | | | | | | | | B5 <= 31: Nivel3 (5.0/1.0)
| | | | | | | | | B5 > 31
| | | | | | | | | | B9 <= 30: Nivel3 (3.0)
| | | | | | | | | | B9 > 30
| | | | | | | | | | | E1 <= 30: Nivel3 (2.0)
| | | | | | | | | | | E1 > 30
| | | | | | | | | | | | E1 <= 35
| | | | | | | | | | | | | B6 <= 23: Nivel3 (2.0)
| | | | | | | | | | | | | B6 > 23: Nivel2 (3.0)
| | | | | | | | | | | | E1 > 35: Nivel2 (3.0)
| | | | | | | | | | B11 > 30: Nivel3 (14.0/1.0)
| | | | | | | B5 > 36
| | | | | | | | E1 <= 29: Nivel3 (4.0/1.0)
| | | | | | | | E1 > 29: Nivel2 (9.0/1.0)
| | | | B8 > 35
| | | | | B1 <= 31: Nivel4 (2.0)
| | | | | E1 > 31: Nivel3 (3.0/1.0)

```

Figura 12. Resultados obtenidos en el WEKA luego de aplicar el algoritmo J48graft a la Base de Hechos.

### 2.4 Árboles de decisión.

Se construye el árbol de decisión a partir de los resultados obtenidos del procesamiento de datos de la Base de Hecho, para lograr con esto un mejor entendimiento del conocimiento que se alcanzó en el procesamiento. Para analizar los resultados se utilizaron varios clasificadores de tipo árbol de decisión. En la tabla a continuación se muestran los resultados obtenidos.

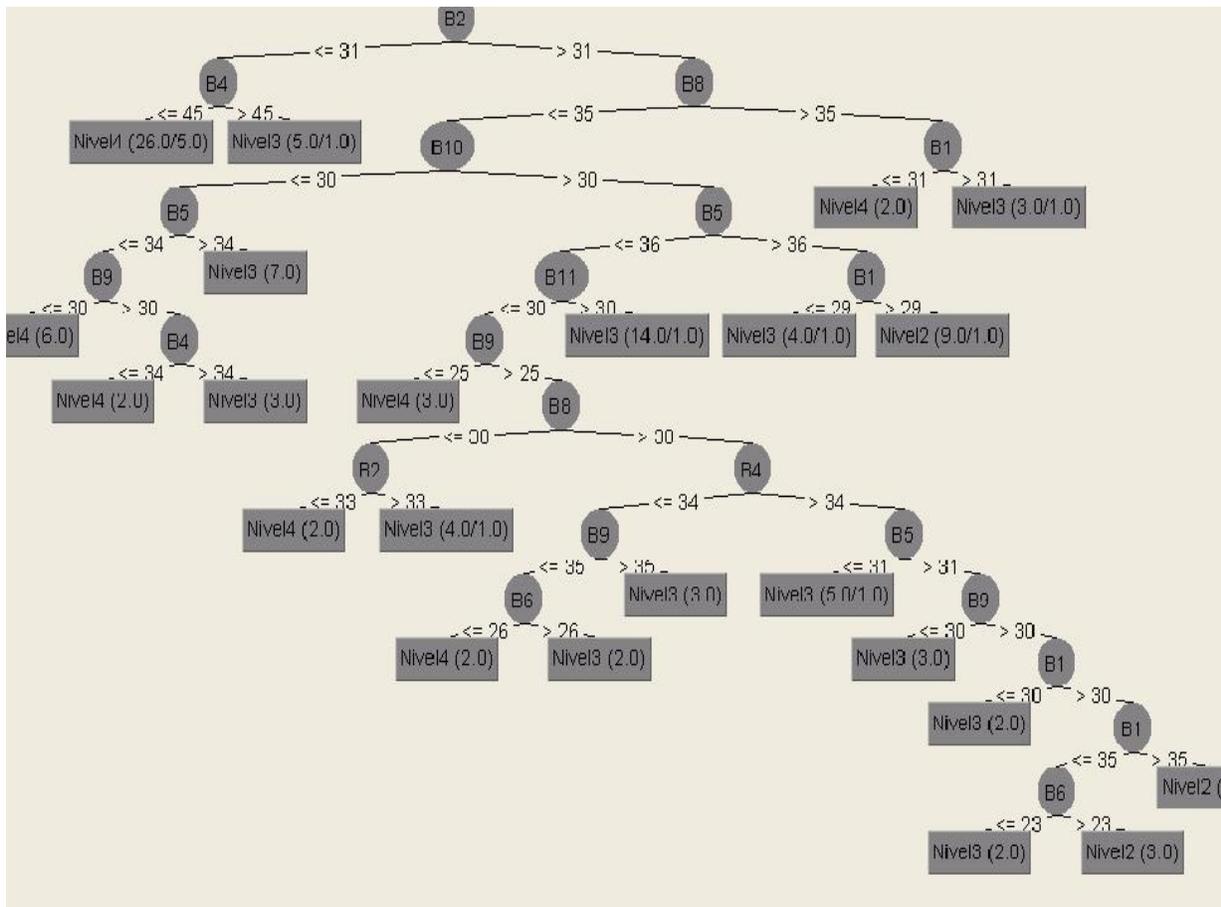
Tipo del clasificador	Calidad de la clasificación	Tiempo de ejecución (s)
J48graft	0.99	0.02
Decision Stump	0.98	0.08

## Capítulo 2: Aplicación del conocimiento

FT	0.99	1.00
Id3	0.97	0.50
J48	0.9	0.08
RandomTree	0.94	0.10

**Tabla 19: Comparación entre algunos clasificadores de tipo árbol.**

Aplicando un clasificador de tipo árbol de decisión J48graft, por ser el de mayor calidad de la clasificación se obtuvo el árbol representado en la en la siguiente figura.



**Figura 13. Árbol de Decisión.**

### Capítulo 3: Resultados Obtenidos

#### Introducción al capítulo

En este capítulo se obtienen un conjunto de reglas que clasifican los objetos de la base de hechos, las cuales constituyen la base de la aplicación. Se mostrará la manera en que fue creado el sistema para obtener la característica de cada estudiante de primer año de la facultad 5, así como las interfaces que lo conforman, sus principales características y algunos resultados obtenidos, luego de realizar algunas pruebas.

#### 3.1 Sistema de Reglas.

En el capítulo anterior durante el procesamiento de datos, se obtuvo un árbol de decisión y la información almacenada en ellos se traducirá en reglas, pudiéndose afirmar que el conocimiento obtenido en el procesamiento se modelará como un sistema de reglas, las cuales constituyen la información esencial que se incluirá en la aplicación y por tanto el conocimiento del cual se apoyará el sistema para dar a los profesores una característica fiable de cada estudiante. En los arboles de decisión el camino hacia cada hoja representa una regla, pudiéndose extraer una amplia variedad de ellas en un mismo árbol, el conocimiento expresado en el árbol se mantiene intacto en el sistema de reglas y la clasificación de ellas resulta ser bastante certera.

##### 3.1.1 Selección del clasificador de tipo regla.

Se generaron un conjunto de reglas aplicando varios clasificadores de tipo reglas. Para la selección del clasificador de tipo regla se analizaron los resultados obtenidos luego de la ejecución de varios clasificadores de este tipo. Utilizando un clasificador de tipo regla NNge, por ser el de mayor calidad de clasificación. En la tabla a continuación se muestran los resultados obtenidos.

## Capítulo 3: Resultados Obtenidos

---

En la siguiente tabla se muestran los resultados obtenidos.

Tipo del clasificador	Calidad de la clasificación	Tiempo de ejecución (s)
NNge	0.99	0.10
Conjuntive hule	0.98	46.70
Decision tablee	0.98	15.08
One R	0.98	1.00
Oidor	0.97	0.30
Error	0.94	10.00

**Tabla 14: Comparación entre algunos clasificadores de tipo reglas.**

Por haberse obtenido la mayor calidad de la clasificación para el clasificador de tipo regla NNge se generaron con el software WEKA las siguientes reglas.

### 3.1.2 Reglas.

Para la selección del clasificador de tipo regla se analizaron los resultados obtenidos luego de la ejecución de varios clasificadores de este tipo. Utilizando un clasificador de tipo regla NNge, por ser el de mayor calidad de clasificación. En la tabla a continuación se muestran los resultados obtenidos.

NNGE classifier

Rules generated:

Class Nivel3 IF:  $26.0 \leq B1 \leq 36.0 \wedge 33.0 \leq B2 \leq 38.0 \wedge 36.0 \leq B3 \leq 43.0 \wedge B4 = 39.0 \wedge 27.0 \leq B5 \leq 38.0 \wedge 8.0 \leq B6 \leq 29.0 \wedge 4.0 \leq B7 \leq 23.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 30.0 \leq B9 \leq 45.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge 30.0 \leq B11 \leq 40.0$  (9)

Class Nivel4 IF:  $B1 = 38.0 \wedge B2 = 30.0 \wedge B3 = 34.0 \wedge B4 = 34.0 \wedge B5 = 28.0 \wedge B6 = 21.0 \wedge B7 = 22.0 \wedge B8 = 35.0 \wedge B9 = 40.0 \wedge B10 = 30.0 \wedge B11 = 30.0$  (1)

## Capítulo 3: Resultados Obtenidos

---

Class Nivel4 IF:  $33.0 \leq B1 \leq 35.0 \wedge 34.0 \leq B2 \leq 36.0 \wedge 37.0 \leq B3 \leq 42.0 \wedge 17.0 \leq B4 \leq 34.0 \wedge B5 = 34.0 \wedge 15.0 \leq B6 \leq 24.0 \wedge 15.0 \leq B7 \leq 19.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge B9 = 30.0 \wedge B10 = 30.0 \wedge B11 = 30.0$  (2)

Class Nivel4 IF:  $33.0 \leq B1 \leq 35.0 \wedge 29.0 \leq B2 \leq 31.0 \wedge 29.0 \leq B3 \leq 31.0 \wedge 45.0 \leq B4 \leq 50.0 \wedge B5 = 32.0 \wedge 4.0 \leq B6 \leq 11.0 \wedge 10.0 \leq B7 \leq 15.0 \wedge B8 = 35.0 \wedge B9 = 30.0 \wedge B10 = 30.0 \wedge B11 = 30.0$  (2)

Class Nivel4 IF:  $13.0 \leq B1 \leq 39.0 \wedge 29.0 \leq B2 \leq 43.0 \wedge 24.0 \leq B3 \leq 44.0 \wedge 22.0 \leq B4 \leq 45.0 \wedge 29.0 \leq B5 \leq 37.0 \wedge 12.0 \leq B6 \leq 24.0 \wedge 13.0 \leq B7 \leq 29.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 20.0 \leq B9 \leq 25.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge B11 = 30.0$  (5)

Class Nivel4 IF:  $25.0 \leq B1 \leq 34.0 \wedge 23.0 \leq B2 \leq 32.0 \wedge 15.0 \leq B3 \leq 40.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 26.0 \leq B5 \leq 35.0 \wedge 11.0 \leq B6 \leq 22.0 \wedge 5.0 \leq B7 \leq 21.0 \wedge B8 = 30.0 \wedge 30.0 \leq B9 \leq 45.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge B11 = 30.0$  (7)

Class Nivel4 IF:  $24.0 \leq B1 \leq 38.0 \wedge 19.0 \leq B2 \leq 32.0 \wedge 33.0 \leq B3 \leq 45.0 \wedge 22.0 \leq B4 \leq 45.0 \wedge 19.0 \leq B5 \leq 27.0 \wedge 14.0 \leq B6 \leq 28.0 \wedge 4.0 \leq B7 \leq 22.0 \wedge B8 = 35.0 \wedge 30.0 \leq B9 \leq 40.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge B11 = 30.0$  (7)

Class Nivel3 IF:  $28.0 \leq B1 \leq 36.0 \wedge 35.0 \leq B2 \leq 42.0 \wedge 30.0 \leq B3 \leq 44.0 \wedge 0.0 \leq B4 \leq 34.0 \wedge 26.0 \leq B5 \leq 41.0 \wedge 10.0 \leq B6 \leq 34.0 \wedge 4.0 \leq B7 \leq 23.0 \wedge 35.0 \leq B8 \leq 40.0 \wedge 35.0 \leq B9 \leq 45.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge 30.0 \leq B11 \leq 35.0$  (10)

Class Nivel2 IF:  $31.0 \leq B1 \leq 37.0 \wedge 34.0 \leq B2 \leq 38.0 \wedge 29.0 \leq B3 \leq 43.0 \wedge 45.0 \leq B4 \leq 50.0 \wedge 35.0 \leq B5 \leq 41.0 \wedge 17.0 \leq B6 \leq 36.0 \wedge 8.0 \leq B7 \leq 18.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 35.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge 30.0 \leq B11 \leq 35.0$  (6)

Class Nivel4 IF:  $B1 = 31.0 \wedge B2 = 38.0 \wedge B3 = 35.0 \wedge B4 = 28.0 \wedge B5 = 33.0 \wedge B6 = 21.0 \wedge B7 = 14.0 \wedge B8 = 30.0 \wedge B9 = 35.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (1)

Class Nivel4 IF:  $29.0 \leq B1 \leq 32.0 \wedge 28.0 \leq B2 \leq 33.0 \wedge 34.0 \leq B3 \leq 35.0 \wedge 39.0 \leq B4 \leq 45.0 \wedge 28.0 \leq B5 \leq 29.0 \wedge 17.0 \leq B6 \leq 27.0 \wedge B7 = 14.0 \wedge B8 = 35.0 \wedge 40.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (2)

Class Nivel4 IF:  $B1 = 38.0 \wedge 33.0 \leq B2 \leq 42.0 \wedge 37.0 \leq B3 \leq 40.0 \wedge 22.0 \leq B4 \leq 28.0 \wedge 15.0 \leq B5 \leq 27.0 \wedge 16.0 \leq B6 \leq 24.0 \wedge 22.0 \leq B7 \leq 27.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 30.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (2)

## Capítulo 3: Resultados Obtenidos

---

Class Nivel4 IF:  $19.0 \leq B1 \leq 23.0 \wedge 38.0 \leq B2 \leq 39.0 \wedge 30.0 \leq B3 \leq 33.0 \wedge 34.0 \leq B4 \leq 39.0 \wedge 26.0 \leq B5 \leq 33.0 \wedge 23.0 \leq B6 \leq 26.0 \wedge 7.0 \leq B7 \leq 11.0 \wedge 35.0 \leq B8 \leq 40.0 \wedge B9 = 35.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (2)

Class Nivel3 IF:  $B1 = 32.0 \wedge B2 = 36.0 \wedge B3 = 36.0 \wedge B4 = 50.0 \wedge B5 = 38.0 \wedge B6 = 3.0 \wedge B7 = 9.0 \wedge B8 = 35.0 \wedge B9 = 30.0 \wedge B10 = 30.0 \wedge B11 = 40.0$  (1)

Class Nivel3 IF:  $30.0 \leq B1 \leq 38.0 \wedge 37.0 \leq B2 \leq 41.0 \wedge 33.0 \leq B3 \leq 35.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 34.0 \leq B5 \leq 36.0 \wedge 11.0 \leq B6 \leq 16.0 \wedge 11.0 \leq B7 \leq 13.0 \wedge B8 = 35.0 \wedge B9 = 35.0 \wedge B10 = 30.0 \wedge B11 = 30.0$  (2)

Class Nivel3 IF:  $33.0 \leq B1 \leq 34.0 \wedge B2 = 33.0 \wedge 23.0 \leq B3 \leq 39.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 35.0 \leq B5 \leq 39.0 \wedge 12.0 \leq B6 \leq 24.0 \wedge 25.0 \leq B7 \leq 31.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 30.0 \leq B9 \leq 45.0 \wedge B10 = 30.0 \wedge 30.0 \leq B11 \leq 35.0$  (2)

Class Nivel3 IF:  $26.0 \leq B1 \leq 36.0 \wedge 27.0 \leq B2 \leq 32.0 \wedge 28.0 \leq B3 \leq 38.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 30.0 \leq B5 \leq 41.0 \wedge 12.0 \leq B6 \leq 36.0 \wedge 5.0 \leq B7 \leq 28.0 \wedge B8 = 35.0 \wedge 30.0 \leq B9 \leq 40.0 \wedge 30.0 \leq B10 \leq 40.0 \wedge 30.0 \leq B11 \leq 35.0$  (6)

Class Nivel4 IF:  $32.0 \leq B1 \leq 34.0 \wedge 24.0 \leq B2 \leq 29.0 \wedge 25.0 \leq B3 \leq 44.0 \wedge 28.0 \leq B4 \leq 34.0 \wedge 32.0 \leq B5 \leq 37.0 \wedge 12.0 \leq B6 \leq 24.0 \wedge 5.0 \leq B7 \leq 20.0 \wedge B8 = 35.0 \wedge 35.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (4)

Class Nivel3 IF:  $B1 = 26.0 \wedge B2 = 38.0 \wedge B3 = 35.0 \wedge B4 = 45.0 \wedge B5 = 36.0 \wedge B6 = 23.0 \wedge B7 = 7.0 \wedge B8 = 35.0 \wedge B9 = 40.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (1)

Class Nivel3 IF:  $B1 = 34.0 \wedge 33.0 \leq B2 \leq 38.0 \wedge B3 = 39.0 \wedge 28.0 \leq B4 \leq 45.0 \wedge 32.0 \leq B5 \leq 34.0 \wedge 14.0 \leq B6 \leq 15.0 \wedge B7 = 3.0 \wedge B8 = 35.0 \wedge B9 = 45.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (2)

Class Nivel3 IF:  $4.0 \leq B1 \leq 32.0 \wedge 36.0 \leq B2 \leq 42.0 \wedge 28.0 \leq B3 \leq 35.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 31.0 \leq B5 \leq 34.0 \wedge 9.0 \leq B6 \leq 30.0 \wedge 10.0 \leq B7 \leq 17.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 35.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge B11 = 35.0$  (4)

Class Nivel2 IF:  $24.0 \leq B1 \leq 36.0 \wedge 39.0 \leq B2 \leq 43.0 \wedge 32.0 \leq B3 \leq 42.0 \wedge 39.0 \leq B4 \leq 50.0 \wedge 32.0 \leq B5 \leq 41.0 \wedge 7.0 \leq B6 \leq 32.0 \wedge 13.0 \leq B7 \leq 30.0 \wedge B8 = 35.0 \wedge 30.0 \leq B9 \leq 45.0 \wedge B10 = 40.0 \wedge B11 = 30.0$  (4)

## Capítulo 3: Resultados Obtenidos

---

Class Nivel2 IF :  $34.0 \leq B1 \leq 37.0 \wedge B2=38.0 \wedge 24.0 \leq B3 \leq 35.0 \wedge B4=39.0 \wedge 34.0 \leq B5 \leq 37.0 \wedge 17.0 \leq B6 \leq 23.0 \wedge 13.0 \leq B7 \leq 16.0 \wedge B8=35.0 \wedge 40.0 \leq B9 \leq 45.0 \wedge B10=40.0 \wedge 30.0 \leq B11 \leq 35.0$  (2)

Class Nivel3 IF :  $B1=28.0 \wedge B2=38.0 \wedge B3=40.0 \wedge B4=45.0 \wedge B5=36.0 \wedge B6=20.0 \wedge B7=27.0 \wedge B8=35.0 \wedge B9=25.0 \wedge B10=40.0 \wedge B11=35.0$  (1)

Class Nivel3 IF :  $26.0 \leq B1 \leq 34.0 \wedge 35.0 \leq B2 \leq 40.0 \wedge 32.0 \leq B3 \leq 37.0 \wedge 34.0 \leq B4 \leq 50.0 \wedge 30.0 \leq B5 \leq 37.0 \wedge 28.0 \leq B6 \leq 34.0 \wedge 8.0 \leq B7 \leq 12.0 \wedge B8=35.0 \wedge B9=25.0 \wedge B10=40.0 \wedge 30.0 \leq B11 \leq 35.0$  (2)

Class Nivel4 IF :  $B1=36.0 \wedge B2=38.0 \wedge B3=28.0 \wedge B4=50.0 \wedge B5=28.0 \wedge B6=29.0 \wedge B7=20.0 \wedge B8=35.0 \wedge B9=25.0 \wedge B10=30.0 \wedge B11=35.0$  (1)

Class Nivel3 IF :  $B1=36.0 \wedge B2=41.0 \wedge B3=40.0 \wedge B4=39.0 \wedge B5=28.0 \wedge B6=26.0 \wedge B7=11.0 \wedge B8=35.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF :  $B1=29.0 \wedge B2=32.0 \wedge B3=43.0 \wedge B4=22.0 \wedge B5=38.0 \wedge B6=20.0 \wedge B7=7.0 \wedge B8=35.0 \wedge B9=45.0 \wedge B10=30.0 \wedge B11=35.0$  (1)

Class Nivel1 IF :  $B1=33.0 \wedge B2=33.0 \wedge B3=44.0 \wedge B4=45.0 \wedge B5=35.0 \wedge B6=36.0 \wedge B7=31.0 \wedge B8=40.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF :  $B1=23.0 \wedge B2=23.0 \wedge B3=39.0 \wedge B4=50.0 \wedge B5=35.0 \wedge B6=11.0 \wedge B7=26.0 \wedge B8=35.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel4 IF :  $B1=31.0 \wedge B2=33.0 \wedge B3=34.0 \wedge B4=28.0 \wedge B5=19.0 \wedge B6=18.0 \wedge B7=18.0 \wedge B8=35.0 \wedge B9=35.0 \wedge B10=30.0 \wedge B11=35.0$  (1)

Class Nivel4 IF :  $B1=31.0 \wedge B2=40.0 \wedge B3=20.0 \wedge B4=34.0 \wedge B5=27.0 \wedge B6=32.0 \wedge B7=12.0 \wedge B8=35.0 \wedge B9=35.0 \wedge B10=30.0 \wedge B11=30.0$  (1)

Class Nivel3 IF :  $B1=34.0 \wedge B2=42.0 \wedge B3=29.0 \wedge B4=45.0 \wedge B5=32.0 \wedge B6=15.0 \wedge B7=8.0 \wedge B8=40.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF :  $B1=35.0 \wedge B2=37.0 \wedge B3=28.0 \wedge B4=45.0 \wedge B5=35.0 \wedge B6=22.0 \wedge B7=19.0 \wedge B8=35.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=35.0$  (1)

Class Nivel3 IF:  $B1=35.0 \wedge B2=37.0 \wedge B3=23.0 \wedge B4=45.0 \wedge B5=33.0 \wedge B6=30.0 \wedge B7=34.0 \wedge B8=30.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

## Capítulo 3: Resultados Obtenidos

---

Class Nivel3 IF:  $33.0 \leq B1 \leq 36.0 \wedge 33.0 \leq B2 \leq 34.0 \wedge B3=36.0 \wedge B4=45.0 \wedge 35.0 \leq B5 \leq 36.0 \wedge 14.0 \leq B6 \leq 16.0 \wedge 7.0 \leq B7 \leq 12.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge 30.0 \leq B9 \leq 35.0 \wedge B10=40.0 \wedge B11=30.0$  (2)

Class Nivel4 IF:  $B1=29.0 \wedge B2=38.0 \wedge B3=35.0 \wedge B4=50.0 \wedge B5=29.0 \wedge B6=26.0 \wedge B7=6.0 \wedge B8=30.0 \wedge B9=30.0 \wedge B10=30.0 \wedge B11=30.0$  (1)

Class Nivel4 IF:  $B1=28.0 \wedge B2=31.0 \wedge B3=27.0 \wedge B4=39.0 \wedge B5=34.0 \wedge B6=21.0 \wedge B7=21.0 \wedge B8=35.0 \wedge B9=30.0 \wedge B10=40.0 \wedge B11=35.0$  (1)

Class Nivel3 IF:  $B1=33.0 \wedge B2=31.0 \wedge B3=30.0 \wedge B4=34.0 \wedge B5=33.0 \wedge B6=23.0 \wedge B7=17.0 \wedge B8=35.0 \wedge B9=30.0 \wedge B10=30.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=36.0 \wedge B2=40.0 \wedge B3=39.0 \wedge B4=50.0 \wedge B5=27.0 \wedge B6=27.0 \wedge B7=29.0 \wedge B8=35.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=35.0 \wedge B2=36.0 \wedge B3=35.0 \wedge B4=39.0 \wedge B5=28.0 \wedge B6=23.0 \wedge B7=4.0 \wedge B8=30.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=35.0$  (1)

Class Nivel2 IF:  $31.0 \leq B1 \leq 34.0 \wedge 32.0 \leq B2 \leq 38.0 \wedge 35.0 \leq B3 \leq 38.0 \wedge 39.0 \leq B4 \leq 45.0 \wedge 36.0 \leq B5 \leq 39.0 \wedge 26.0 \leq B6 \leq 27.0 \wedge 1.0 \leq B7 \leq 3.0 \wedge 30.0 \leq B8 \leq 35.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=30.0$  (2)

Class Nivel2 IF:  $B1=37.0 \wedge B2=40.0 \wedge B3=33.0 \wedge B4=34.0 \wedge B5=37.0 \wedge B6=23.0 \wedge B7=19.0 \wedge B8=35.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=35.0$  (1)

Class Nivel2 IF:  $B1=30.0 \wedge B2=38.0 \wedge B3=35.0 \wedge B4=45.0 \wedge B5=38.0 \wedge B6=15.0 \wedge B7=20.0 \wedge B8=35.0 \wedge B9=5.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=36.0 \wedge B2=33.0 \wedge B3=36.0 \wedge B4=45.0 \wedge B5=31.0 \wedge B6=15.0 \wedge B7=20.0 \wedge B8=30.0 \wedge B9=25.0 \wedge B10=40.0 \wedge B11=40.0$  (1)

Class Nivel3 IF:  $B1=31.0 \wedge B2=27.0 \wedge B3=41.0 \wedge B4=50.0 \wedge B5=37.0 \wedge B6=28.0 \wedge B7=3.0 \wedge B8=30.0 \wedge B9=35.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=36.0 \wedge B2=37.0 \wedge B3=27.0 \wedge B4=45.0 \wedge B5=36.0 \wedge B6=24.0 \wedge B7=8.0 \wedge B8=35.0 \wedge B9=30.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel4 IF:  $B1=29.0 \wedge B2=43.0 \wedge B3=16.0 \wedge B4=45.0 \wedge B5=34.0 \wedge B6=24.0 \wedge B7=13.0 \wedge B8=40.0 \wedge B9=40.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=36.0 \wedge B2=31.0 \wedge B3=34.0 \wedge B4=45.0 \wedge B5=27.0 \wedge B6=27.0 \wedge B7=6.0$   
 $\wedge B8=35.0 \wedge B9=45.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel3 IF:  $B1=39.0 \wedge B2=32.0 \wedge B3=44.0 \wedge B4=39.0 \wedge B5=27.0 \wedge B6=18.0 \wedge$   
 $B7=11.0 \wedge B8=35.0 \wedge B9=40.0 \wedge B10=40.0 \wedge B11=30.0$  (1)

Class Nivel5 IF:  $B1=25.0 \wedge B2=25.0 \wedge B3=45.0 \wedge B4=28.0 \wedge B5=30.0 \wedge B6=9.0 \wedge B7=9.0$   
 $\wedge B8=35.0 \wedge B9=20.0 \wedge B10=30.0 \wedge B11=30.0$  (1)

### **3.2 Solución al problema planteado.**

#### **3.2.1 Características de la aplicación.**

Para dar solución al problema que dio origen a este trabajo, se creó una aplicación cuyas principales características son las siguientes:

- Creada en el lenguaje C++.
- Programación orientada a objetos.
- Permite buscar las características de un estudiante a partir de un identificador (CI).
- Permite a un administrador llenar la base de datos siempre que tenga acceso a la opción llenar datos del estudiante.
- Se clasifica como un sistema basado en reglas. Las reglas que lo soportan son las obtenidas por el software Weka.

#### **3.2.2 Interfaces de la Aplicación.**

Se creó una sección administrador, el cual es el responsable de adicionar los datos personales de cada estudiante y los valores obtenidos en cada una de las secciones de las encuestas.

## Capítulo 3: Resultados Obtenidos

---

La figura a continuación muestra la interfaz administrador, donde el usuario y la contraseña deben ser correctos para que poder acceder al formulario de adicionar estudiante y adicionar encuesta.



**Figura 14. Formulario de Acceso del Administrador.**

Si el administrador de la aplicación no introduce el nombre de usuario o contraseña correctamente la aplicación se encargará de mostrar un mensaje de error, para que introduzca los datos correctamente, la figura a continuación es una muestra de error al acceder a la aplicación.



**Figura 15. Formulario de error al acceder el administrador a la aplicación.**

## Capítulo 3: Resultados Obtenidos

Luego que el administrador accede correctamente a la aplicación tiene permiso total para adicionar los estudiantes a la base de datos y los valores obtenidos en las secciones de las encuestas, esta operación se hace manualmente y a través de los siguientes formularios.

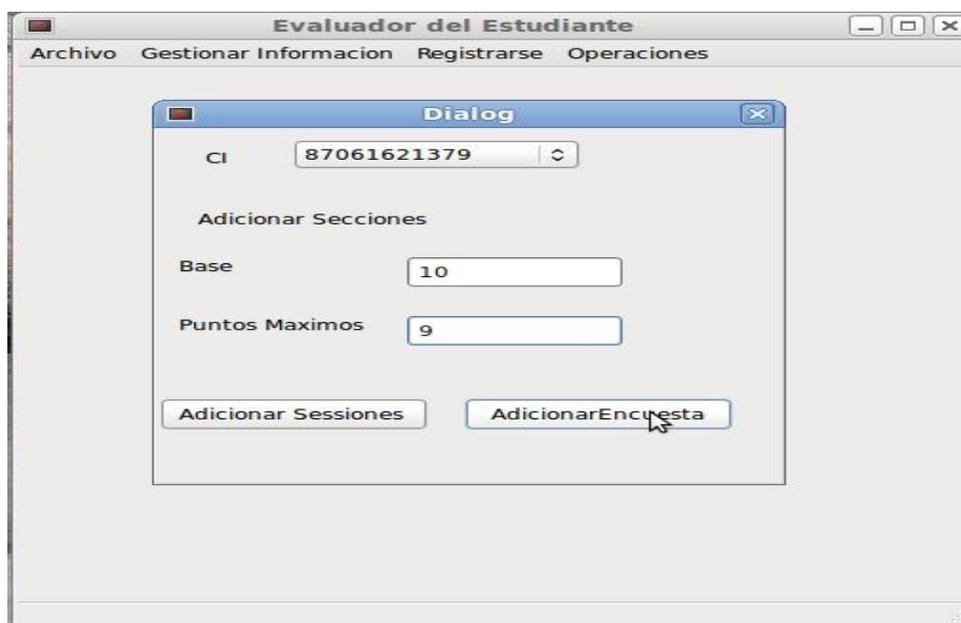


The screenshot shows a window titled 'Evaluador del Estudiante' with a menu bar containing 'Archivo', 'Gestionar Informacion', 'Registrarse', and 'Operaciones'. A sub-window titled 'Adicionar Estudiantes' is open, containing the following fields:

- Nombre: Orlys
- Apellido: Valero
- CI: 87061621379
- Grupo: 5101
- Situación Económica: Bien
- Procedencia Estudiante: IPVCE
- Zona: Rural
- Escolaridad Padre: Universitario
- Escolaridad Madre: Universitario

At the bottom of the sub-window are two buttons: 'Aceptar' and 'Cancelar'.

Figura 16. Formulario adicionar estudiante.



The screenshot shows the same 'Evaluador del Estudiante' window. A sub-window titled 'Dialog' is open, containing the following fields and buttons:

- CI: 87061621379
- Adicionar Secciones
- Base: 10
- Puntos Maximos: 9
- Buttons: 'Adicionar Sesiones' and 'Adicionar Encuesta' (with a mouse cursor over it)

Figura 17. Formulario adicionar encuesta.

## Capítulo 3: Resultados Obtenidos

Después que los datos los estudiantes están almacenados en la base de datos el usuario puede acceder al formulario de caracterizar estudiante para conocer las características del mismo. A continuación se muestra el formulario de caracterizar estudiante.



**Figura 18. Formulario caracterizar estudiante.**

Como resultado de dar clic en caracterizar se obtiene el mensaje que a continuación se muestra en la figura 19, donde se obtiene el nivel del estudiante.



**Figura 19. Formulario caracterizar estudiante.**

### **Conclusiones**

En esta investigación fueron procesados los resultados de los diagnósticos aplicados a los estudiantes de primer año de la Facultad 5 de la Universidad de las Ciencias Informáticas. La muestra toma fue de 115 estudiantes representando el 70% de la matrícula. Se analizaron cada uno de los 7 ejercicios comprendidos en el diagnóstico, obteniendo como resultado final la matriz de decisión, donde las filas representan los objetos y las columnas los atributos incluyendo el de decisión. A partir de esta respuesta se generaron las reglas de inferencia mediante el software Weka y a su vez se generó el árbol de decisión utilizando un clasificador de tipo Trees. Estas reglas de producción fueron programadas en QT, librería para la creación de interfaces gráficos, orientado a objetos, lo que facilita el desarrollo de software. El lenguaje para el que se encuentra disponible es C++, es multi-plataforma y de código abierto. Se cumplió el objetivo de la investigación desarrollando una aplicación que permite determinar las principales características cognoscitivas de los estudiantes que ingresan en el primer año en la facultad 5 de la UCI, dando así respuesta al problema planteado.

### **Recomendaciones**

- Utilizar los resultados de esta aplicación en el próximo curso.
- Extender este tipo de investigación a toda la UCI.
- Realizar un proceso similar para los estudiantes de tercer año, con el objetivo de que el rol asignado en los proyectos productivos este en correspondencia con las capacidades cognoscitivas de los mismos.

### Referencias bibliográficas

1. Algunas definiciones de Inteligencia Artificial [Citado: febrero 16, 2010.]  
<http://www.infor.uva.es/~calonso/IAI/Tema1-introduccionIAI/DefinicionesIA.pdf>
2. Béjar, Javier. Apuntes de Aprendizaje Automático. Curso 2005/2006.
3. Z. Pawlak. "Rough sets". International Journal of Information & Computer Sciences. Vol. 11. 1982. pp. 341-356.
4. J. Komorowski, Z. Pawlak, Z. "Rough Sets: A tutorial. In Pal, S.K. and Skowron, A. (Eds.) Rough Fuzzy Hybridization: A new trend in decision-making." Springer. 1999. pp. 3-98.
5. Pérez, Rafael Bello. La teoría de conjuntos aproximados en el contexto de la inteligencia artificial. Universidad de las Ciencias Informáticas. Cuba. 2008.
6. Ing.González Remón Neyvis. Trabajo de diploma Principales rasgos para la predicción del éxito o fracaso de estudiantes de la Universidad de las Ciencias Informáticas mediante la teoría de conjuntos aproximados de la Ing.Neyvis Remón González.
7. Carrillo, Fernando Virseda Benito y Javier Román. Minería de datos y aplicaciones. [Disponible] Universidad CarlosIII. [Citado: marzo 13, 2009.]  
<http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>.
8. 2007 Yailé Caballero / Delia Álvarez / Analay Baltá / Rafael Bello / María García. Un nuevo algoritmo de selección de rasgos basado en la Teoría de los Conjuntos Aproximados. Revista Facultad de Ingeniería Universidad de Antioquia, septiembre, número 041 Universidad de Antioquia Medellín, Colombia pp. 132-144.  
<http://redalyc.uaemex.mx>.
9. Langley. Selection of Relevant Features in Machine Learning. Procs. Of the AAAI Fall Symposium on Relevance. New Orleans, LA : s.n., 2004.
10. Bahamonde, Antonio. [Disponible] [Citado: marzo 26, 2009.] [www.aic.uniovi.es](http://www.aic.uniovi.es).
11. Diagnóstico financiero. [Citado: febrero 13, 2010.][http://www.angelfire.com/ex2/john\\_rosso/Posgrados/A\\_FIN.pdf](http://www.angelfire.com/ex2/john_rosso/Posgrados/A_FIN.pdf).
12. Corzo, Yuliana. La Lógica Difusa, desde Porlamar, Venezuela. <mailto:yulicorzo@cantv.net>.

## Referencias Bibliográficas

---

13. Morate, Diego García. WEKA. España: [Disponible] 2005 [Citado: enero 21, 2009.] <http://metaemotion.com/diego.garcia.morate/download/WEKA.pdf>.
14. Kirkby, Richard; Frank, Eibe; Reutemann, Peter. WEKA Explorer User Guide for Versión 3-5 [Disponible] 2008 [Citado: enero 22, 2009.] <http://www.cs.waikato.ac.nz/~ml/WEKA/>.
15. Curso básico de Statgraphics [Citado: febrero 17, 2010.] <http://www.sisoft.ucm.es/Manuales/Statgraphics7.pdf>
16. Eneko Astigarraga, Método Delphi. Facultad de CC.EE. y Empresariales. ESTE
17. Mundaiz, 50 Apartado 1.359 E-20.080 Donostia - San Sebastián. Universidad de
18. Deusto. Disponible en: [http://www.unalmed.edu.co/~poboyca/documentos/documentos1/documentosJuan%20Diego/Plnaifi\\_Cuencas\\_Pregrado/Sept\\_29/Metodo\\_delphi.pdf](http://www.unalmed.edu.co/~poboyca/documentos/documentos1/documentosJuan%20Diego/Plnaifi_Cuencas_Pregrado/Sept_29/Metodo_delphi.pdf).
19. Bravo, M. de L. y Arrieta, J. J.: Revista Iberoamericana de Educación (ISSN: 1681-5653). El método Delphi. Su implementación en una estrategia didáctica. Disponible en:
20. <http://www.rieoei.org/deloslectores/804Bravo.PDF>.
21. KONOW, I. y PÉREZ, G. Método Delphi. [Disponible] 1998. [Citado: marzo 23, 2009.] <http://geocities.com/Pentagon/Quarters/7578/pros01-03.html>.

### Bibliografía

- Amézquita, Marcela Quintero Edgar. *Herramienta "Arboles de Decisión": Alternativas de Uso de la Tierra para los Llanos Orientales de Colombia*. [Disponible] 2000. [Citado: marzo 16, 2009.] [http://ciatlibrary.ciat.cgiar.org/documentos\\_electronicos\\_ciat/articulos\\_ciat/Manual\\_Arboles.pdf](http://ciatlibrary.ciat.cgiar.org/documentos_electronicos_ciat/articulos_ciat/Manual_Arboles.pdf).
- Bahamonde, Antonio. [Disponible] [Citado: marzo 26, 2009.] [www.aic.uniovi.es](http://www.aic.uniovi.es).
- Bravo Estévez, María de Lourdes; Arrieta Gallastegui, José Joaquín. *El método Delphi como estrategia didáctica para la enseñanza universitaria*, Universidad de Cienfuegos. Cuba. Universidad de Oviedo. España. [Disponible] 2006. [Citado: marzo 23, 2009.] <http://www.rieoei.org/deloslectores/804Bravo.PDF>
- Caballero, Yailé; Bello, Rafael; Álvarez, Delia; García, María M.; Baltá, Analay. *Un nuevo algoritmo de selección de rasgos basado en la Teoría de los Conjuntos Aproximados*. [Disponible] 2007 [Citado: febrero 17, 2009.] <http://redalyc.uaemex.mx/redalyc/pdf/430/43004111.pdf>
- Carrillo, Fernando Virseda Benito y Javier Román. *Minería de datos y aplicaciones*. [Disponible] Universidad Carlos III. [Citado: marzo 13, 2009.] <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>.
- D.A SAVIC, J.W DAVIDSON, DAVIS. *Data Mining and Knowledge discovery for the water industry*, *Water Industry Systems, modelling and optimisation applications*. s.l. : Dragan A. Savic and Godfrey A. Walter, 2001.
- Fernández Caballero, Antonio; García, María; Arjona, Manzano. *Una Perspectiva de la Inteligencia Artificial en su 50 Aniversario*. [Disponible] Julio 14, 2006 [Citado: febrero 19, 2009.] <http://www.infob.uclm.es/personal/AntonioFdez/download/papers/conference/cmpi2006-volumell.pdf#page=219>.
- García Corzo, Pablo M. *Ciencia Abierta y Software Libre en el aprendizaje* [Disponible] 19 de noviembre de 2008 [Citado: abril 20, 2009.] <http://www.ucm.es/info/aulasun/archivos/presenta.pdf>.

- KONOW, I. y PÉREZ, G. *Método Delphi*. [Disponible] 1998. [Citado: marzo 23, 2009.] <http://geocities.com/Pentagon/Quarters/7578/pros01-03.html>.
- Kirkby, Richard; Frank, Eibe; Reutemann, Peter. *WEKA Explorer User Guide for Versión 3-5-8*. [Disponible] 2008 [Citado: enero 22, 2009.] <http://www.cs.waikato.ac.nz/~ml/WEKA/>.
- Langley. *Selection of Relevant Features in Machine Learning. Procs. Of the AAAI Fall Symposium on Relevance*. New Orleans, LA : s.n., 2004.
- MARAKAS, G. *Decision Support Systems in the 21st Century*. New York, E.U.A. : Prentice-Hall, 1998.
- Morate, Diego García. *WEKA*. España : [Disponible] 2005 [Citado: enero 21, 2009.] <http://metaemotion.com/diego.garcia.morate/download/WEKA.pdf>.
- Murillo, Bravo. *Integración del STATGRAPHICS en un programa Seis Sigma*. [Disponible] 2004. [Citado: abril 6, 2009.] <http://www.STATGRAPHICS.net/SeisSigma.pdf>.
- Orlowska, E. (ed.). *Incomplete Information: Rough set analysis*. Physica- Verlag : s.n., 1998.
- PARISCA, S. *Método Delphi. Gestión tecnológica y competitividad Estrategia y filosofía para alcanzar la calidad total y el éxito en la gestión impresional*. La Habana 1997.
- Pawlak, Z. *Rough Sets. International journal of Computer and Information Sciences*. 1982.
- PAWLAK, Z. *Rough Sets, International Journal of Information & Computer Sciences*. 1982. vol. 11, p. 341-356.
- Pawlak, Z. *Rough sets*. 1995.
- Pawlak, Z. *Vagueness and uncertainty: a rough set perspective en Computational Intelligence*. 1995.
- PAWLAK, Z. *Rough Sets – Theoretical Aspects of Reasoning about Data* . Boston, London : Kluwer Academic Publishers, 2001.

### Capítulo 3. Resultados obtenidos

---

- PAWLAK, Z. and SLOWINSKI, R. *Decision Analysis using Rough Sets. ICS Research Report no 21.* Warsaw, Poland : Institute of Computer Science, Warsaw University of Technology, 1993.
- Pérez, Rafael Bello. *La teoría de conjuntos aproximados en el contexto de la inteligencia artificial.* Universidad de las Ciencias Informáticas. Cuba. 2008.
- RUIZ OLABUÉNAGA, J. e ISPIZUA, M. A. La técnica Delphi. *La descodificación de la vida cotidiana. Métodos de investigación cualitativa* 1989.
- Vidal, Tomás Arredondo. *Árboles de Decisión* [Disponible] marzo 26, 2008 [Citado: marzo 16, 2009.] <http://profesores.elo.utfsm.cl/~tarredondo/info/soft-omp/Arboles%20de%20Decision.pdf>
- Zhang S., Zhang C. , Yang Q. *Data preparation for data mining. Applied Artificial Intelligence.* 2005.