

Universidad de las Ciencias Informáticas

Facultad 5



**Propuesta de análisis de costo del proyecto SCADA
mediante la teoría de los conjuntos aproximados.**

Trabajo de diploma para optar por el título de
Ingeniero en Ciencias Informáticas.

Autores: Yenismara Pelayo García.
María Evelia Torres Batista.

Tutor: MSc.Roberto Millet Luaces.

Co-tutora: Ing. Neyvis Remón González.

Asesores: MSc. Ana Rita Poyeaux Vidal.
Lic. Adrián Garrido Hernández

Ciudad de La Habana, Cuba

Junio, 2010

"Año 52 de la Revolución"

*” La medición es el primer paso para el control y la mejora.
Si no se puede medir, no se puede entender.
Si no se entiende, no se puede controlar.
Si no se puede controlar, no se puede mejorar.”*

H. James Harrington.

DECLARACIÓN DE AUTORÍA

Declaramos ser los únicos autores de este trabajo y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Firma de Autor
(María Evelia Torres Batista)

Firma de Autor
(Yenismara Pelayo García)

Firma del Tutor
(MSc Roberto Millet Luaces)

Firma de la Co-tutora
(Ing. Neyvis Remón González.)

DATOS DEL CONTACTO

Síntesis curricular del Tutor:

Roberto Millet Luaces.

Especialidad: Ingeniero Eléctrico en 1986, en la Universidad de Camagüey.

Categoría Docente: Profesor Auxiliar

Categoría Científica: MSc. (Máster en Ciencias de la Matemática.).

Imparte docencia en universidades desde 1987.

Tesis tutoradas: 10 en Ingeniería Eléctrica, 9 en Informática.

Años de graduado: 25

Imparte docencia en universidades desde 1987.

Ubicación: Universidad de las Ciencias Informáticas, Cuba.

E-mail: milletp@uci.cu

Síntesis curricular de la Co-tutora:

Neyvis Remón González

Especialidad: Ingeniera en Ciencias Informáticas en el año 2009.

Categoría Docente: Instructor recién graduado.

Ubicación: Universidad de las Ciencias Informáticas, Cuba.

DEDICATORIA

A mis padres Odalis Batista Pérez y José Luis Torres Góngora que confiaron siempre en mí, me han amado y enseñado a seguir el camino correcto, quienes dieron con amor todo cuanto pudieron para que lograra este éxito de hoy.

A mi abuela Evelia por todo el apoyo, preocupación constante, la confianza y el amor que siempre me ha dado en la vida. Sin su ejemplo no hubiera podido lograr lo que he sido hasta hoy.

A mí querido abuelo Alternán que aunque ya no este conmigo siempre fue un guía incondicional y estará en mi corazón.

A mi familia que se ha preocupado por mí en cada momento y han ayudado de una manera u otra a conquistar este logro.

A mis amigos por escuchar mis pensamientos, comprenderme, convivir con mis defectos, ayudarme en los buenos y en los malos momentos; y sobre todas las cosas por confiar en mí.

A todos gracias por ayudarme a lograrlo.

Mary

Desde lo más profundo de mi corazón le dedico este trabajo a las 2 personas más importantes para mí y que más amo en este mundo: A mi madre Sonia y Abuela Mirna porque gracias a su ayuda, amor y dedicación pude convertir este sueño en realidad.

Yeny

AGRADECIMIENTOS

A mis padres por haberme apoyado en todo momento, dando lo mejor de sí para que saliera adelante en mi carrera universitaria, con el fin de convertirme en una profesional; nunca podré pagarles todo lo que han hecho por mí.

A toda mi familia por su dedicación, sacrificio, preocupación constante y el apoyo a lo largo de mi vida escolar.

A mi amiga Yailin por brindarme su ayuda en todo lo que me ha hecho falta.

A mis amigos por ayudarme continuamente, confiar en mí y haber estado presentes en las buenas y las malas.

Al tutor por depositar confianza en nosotras, brindarnos su apoyo diario y estar presente en todo momento de nuestro trabajo de diploma.

A la cotutora por el tiempo que nos dedicó y su deseo de vernos triunfar.

A mi compañera de tesis por el esfuerzo realizado para lograr que este sueño se hiciera realidad.

A Ana Rita Poyeaux, Adrián Garrido, Juan Antonio Plasencia, Luis Enrique García y a otros especialistas por su cooperación en la elaboración de esta investigación.

Al tribunal por el apoyo y la colaboración en el perfeccionamiento de este trabajo.

A todos los profesores por contribuir con su ejemplo en la formación de mi educación.

A los que estuvieron a mi lado cuando me hizo falta e hicieron posible la realización de este trabajo.

A la Revolución cubana y a Fidel por sus enseñanzas, la oportunidad de formarme en esta Universidad. Gracias a ello hoy soy profesional.

Mary

A mis padres (Sonia y Antonio) por el amor, comprensión y dedicación que me han dado. Por haber hecho de mí la persona que hoy en día soy y sobre todo, por haberme ayudado a cumplir todos mis sueños en esta vida. Para ellos todo mi amor y agradecimientos.

A mi abuelita Mirna, por no faltar nunca cuando la necesité, por complacer todos mis caprichos y tener tanta paciencia. Te quiero mucho.

A mi abuelito Beto, por quererme siempre y preocuparte por mí en todo momento.

A mi abuelita Cuka, por su cariño, su amor y por brindarme momentos de alegría.

A mi hermanito Yoel, por apoyarme en los momentos más difíciles. Te quiero mucho.

A mis sobrinas, por darme tanta alegría y cariño.

A toda mi familia en especial mis tíos (Robertico y Luisi) y mis primos (Roberto Carlos, Yamilé, Ronileyxi y Robert Yoel) que se han preocupado por mí en estos cinco años de sacrificio.

A mis mejores amigas (Dalianny y Elizabeth) gracias por sus consejos y las largas horas de charlas, por compartir buenos y malos momentos, por estar siempre ahí para mí.

A mi amigo Guillermito, por su cariño y haberme dado fuerzas para optar por este sueño.

A mi compañera de tesis María, por tanto esfuerzo que hicimos juntas para realizar nuestro sueño.

A mis tutores (Millet y Neyvis) por la ayuda brindada incondicional en este tiempo.

A Ana Rita Poyeaux, Adrián Garrido, Juan Antonio Plasencia, Luis Enrique García y a otros especialistas por su cooperación en la elaboración de esta investigación.

A todos compañeros del centro de trabajo de mi mamá, por su apoyo y cariño en especial (Pablo, Chuchi, Beoto, Correa y María).

A todos mis compañeros de cinco años (de aulas, de apartamento y de recrees), gracias por conocerlos...

Yeny

RESUMEN

En esta investigación se utiliza la teoría de los conjuntos aproximados, un método creado en 1982 por Z. Pawlak, que proporciona un conjunto de herramientas que se pueden aplicar en problemas de minería de datos y descubrimiento del conocimiento, entre las cuales se encuentra el análisis de costo. Se tratan conceptos relacionados con las técnicas de Inteligencia Artificial, la teoría de los conjuntos aproximados y el análisis de costo. Se utilizan elementos del método Delphi de la consulta a expertos con la que se valida la entrada de datos a utilizar. Se elabora una Base de Hechos a partir de varias plantillas oficiales empleadas en la UCI y de los gastos en los que incurre el proyecto productivo SCADA. Se realiza la obtención del conocimiento de la Base de Hechos utilizando el software WEKA del que se derivó un conjunto de reglas de producción, que constituyen el soporte de la aplicación realizada, que permite conocer si un proyecto es rentable o no.

PALABRAS CLAVES: teoría de conjuntos aproximados, análisis de costo, inteligencia artificial.

ÍNDICE DE CONTENIDO

DEDICATORIA	V
AGRADECIMIENTOS.....	VII
RESUMEN.....	IX
INTRODUCCIÓN.....	1
1. CAPÍTULO I FUNDAMENTACIÓN TEÓRICA	6
1.1 Introducción.....	6
1.2 Análisis de costo.....	6
1.3 Conceptos relacionados con Análisis de costo.....	7
1.3.1 Costo.....	7
1.3.2 Sistemas de costos.....	8
1.3.3 Presupuesto de costo.....	8
1.3.3.1 Importancia de los Costos.....	8
1.3.3.2 La estimación del costo de un proyecto.....	9
1.3.3.3 Costos de producción.....	9
1.3.3.3.1 Clasificación de los costos de producción.....	10
1.3.4 Solvencia.....	11
1.3.5 Precio de venta.....	11
1.3.6 Utilidad o pérdida.....	12
1.3.7 Margen de utilidad.....	12
1.3.8 Toma de decisión.....	12
1.3.9 Rentabilidad del producto.....	13
1.4 Inteligencia Artificial.....	13
1.4.1 Algunos conceptos relacionados con Inteligencia Artificial.....	14
1.4.1.1 Aprendizaje automático.....	15
1.4.1.1.1 Algoritmos de aprendizaje automático.....	15
1.4.1.2 Minería de datos.....	18
1.4.1.3 Selección de rasgos.....	18
1.4.2 Teoría de los conjuntos aproximados.....	18
1.4.2.1 Sistema de Información.....	19
1.4.2.2 Sistema de decisión.....	20
1.4.2.3 Aproximaciones de un conjunto.....	21

1.4.2.4	Algunas propiedades de los conjuntos aproximados.	22
1.4.2.5	Medidas asociadas a conjuntos aproximados.	23
1.4.2.6	Reglas de decisión.	24
1.4.2.7	¿Cómo podemos usar RST para tomar decisiones?.....	25
1.4.2.8	Ventajas de esta teoría en comparación a otras técnicas de la IA.....	25
1.5	Descripción de las herramientas y técnicas a utilizar.	27
1.5.1	Método Delphi de la consulta de expertos.	27
1.5.2	Statgraphics.....	27
1.5.3	WEKA.....	28
1.5.4	ROSE2.	29
1.5.5	Lenguaje de programación.	29
2.	CAPÍTULO II APLICACIÓN DEL CONOCIMIENTO	30
2.1	Introducción	30
2.2	Método Delphi de la consulta de expertos.....	30
2.2.1	Cuestionario aplicado a los Expertos.....	30
2.2.1.1	Cuestionario de preguntas.....	30
2.2.1.2	Resultados obtenidos.	32
2.3	Análisis estadístico de la información.....	36
2.3.1	Gráficos Obtenidos.....	38
2.4	Elaboración de la Base de Hechos.	40
2.4.1	Propuesta del Departamento de Ciencias Empresariales.	40
2.4.2	Plantilla proporcionada por ALBET.....	41
2.4.3	Modelo a utilizar.....	42
2.4.3.1	Descripción de algunos elementos que contiene el modelo.....	43
2.5	Base de hechos.....	52
2.6	Sistema de decisión.....	53
3.	CAPÍTULO III. PROPUESTA DE SOLUCIÓN.....	56
3.1	Introducción.....	56
3.2	Procesamiento de datos.	56
3.2.1	Preparación de los Datos en el ROSE.	56
3.2.1.1	Procesamiento de la Base de Hechos y resultados obtenidos.....	57
3.2.2	Preparación de los Datos en el WEKA.....	59
3.2.2.1	Procesamiento de la Base de Hechos y resultados obtenidos.....	60

3.3	Árbol de decisión.....	60
3.3.1	Pruebas.....	64
3.3.1.1	Resultados obtenidos en las pruebas.....	64
3.4	Sistema de Reglas.....	66
3.4.1	Reglas generadas por el WEKA.....	66
3.5	Características de la aplicación.....	69
3.6	Interfaces de la aplicación.....	69
3.7	Pruebas realizadas a la aplicación.....	72
4.	CONCLUSIONES	73
5.	RECOMENDACIONES	74
6.	ANEXOS.....	75
	Anexo 1.....	75
	Anexo 2.....	76
	Anexo 3.....	77
	Anexo 4.....	78
	CITAS BIBLIOGRÁFICAS.....	79
	BIBLIOGRAFÍA.....	84
	GLOSARIO DE TÉRMINOS	90

ÍNDICE DE TABLA

Tabla 1: Representación de un sistema de información.	20
Tabla 2: Cuestionario de preguntas.....	31
Tabla 3: Resultados obtenidos.	33
Tabla 4: Identificador de cada variable.....	33
Tabla 5: Matriz Booleana.....	35
Tabla 6: Tabla de poderes.	36
Tabla 7: Análisis estadístico entre las variables.....	37
Tabla 8: Descripción de los elementos que contiene el modelo.....	52
Tabla 9: Identificador para cada atributo.....	53
Tabla 10: Valores de la Base de Hechos.....	54
Tabla 11: Ejemplo de la representación del sistema de decisión.	55
Tabla 12: Comparación entre algunos clasificadores de tipo árbol.	60
Tabla 13: Comparación entre algunos clasificadores de tipo reglas.	66

ÍNDICE DE FIGURAS

Figura 1: Clasificación de los costos de producción.	10
Figura 2: Toma de decisiones.	13
Figura 3: La teoría de conjuntos aproximados en el contexto de la Inteligencia Artificial.	19
Figura 4: Grafo de relaciones entre las variables.	34
Figura 5: Gráfico de la relación de las variables CA-FC.....	38
Figura 6: Gráfico de la relación de las variables GC-TD.....	38
Figura 7: Gráfico de la relación de las variables TD-AD.	39
Figura 8: Gráfico de la relación de las variables CP-RP.	39
Figura 9: Gráfico de la cantidad de frecuencia con que los expertos respondieron.	39
Figura 10: Representación del por ciento de la concordancia de los expertos.	40
Figura 11: Plantilla para el control de los elementos y subelementos de gastos.....	41
Figura 12: Plantilla proporcionada por ALBET.	42
Figura 13: Modelo a utilizar.	43
Figura 14: Archivos .isf.....	57
Figura 15: Aproximaciones.....	58
Figura 16: Archivos .arff	59
Figura 17: Árbol de Decisión con ADTree.	63
Figura 18: Interfaz de inicio.....	70
Figura 19: Interfaz de Analizador.....	71
Figura 20: Interfaz de Ayuda.	71
Figura 21: Base de Hechos en Excel.	75
Figura 22: Procesamiento de la Base de Hechos en WEKA.	76
Figura 23: Clasificación en WEKA.....	77
Figura 24: Representación del sistema de decisión en WEKA.	78

INTRODUCCIÓN

La Inteligencia Artificial (IA) es una rama de la Ciencia de la Computación, que se utiliza para la creación de hardware y software que imitan el pensamiento humano. Su principal objetivo es llevar a la computadora las amplias capacidades del pensamiento y, para esto las convierte en entes inteligentes con la creación de software. Lo que les permite imitar algunas de las funciones del cerebro humano en aplicaciones particulares. El fin no es reemplazar al hombre, sino proveerlo de una herramienta poderosa para asistirlo en su trabajo.

En la IA, se han desarrollado nuevas herramientas denominadas “inteligencia computacional”, las cuales tienen aplicación en la gestión financiera empresarial especialmente en el tratamiento de la información imprecisa, dentro de estas técnicas se puede mencionar: la teoría de los conjuntos aproximados (*rough sets theory*, en inglés).

Algunas investigaciones sobre la teoría de los conjuntos aproximados se han realizado en varias partes del mundo, como por ejemplo en la Universidad Complutense de Madrid sobre "Predicción de insolvencias con el método Rough Set"[9], “La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multiatributo” [5], “La metodología Rough Set frente al análisis discriminante en la predicción de insolvencia en empresas aseguradoras” [17], de los investigadores M.J. Segovia Vargas, J.A. Gil Fana, A. Heras Martínez, J.L. Vilar Zanón y “Predicción de crisis empresariales en seguros no vida mediante la metodología Rough Set” de la investigadora María Jesús Segovia Vargas. [15]

En Cuba se han desarrollado investigaciones sobre esta teoría; en la Universidad de Camagüey y la Universidad Central Marta Abreu de Las Villas por Rafael Bello, Yailé Caballero Mota y otros investigadores como “Selección de rasgos relevantes para el análisis de datos a través de los conjuntos aproximados.”[10] y “Un nuevo algoritmo de selección de rasgo basado en la teoría de conjuntos aproximados”. [14]

Con el avance de la tecnología en el ámbito empresarial se han desarrollado sistemas de IA que facilitan el trabajo en las decisiones económicas, ya sea como soporte para la toma de decisiones, ayuda de labores, tareas que exijan gran demanda de tiempo o representen un alto grado de peligrosidad al ser

humano. Un ejemplo de esta aplicación son los sistemas informáticos para la economía. Muchas investigaciones son las que se han realizado en el mundo sobre este tema, entre las que se puede mencionar en la Universidad del Valle (Colombia) sobre “Análisis de costo en sistemas de potabilización de agua” por Hernán A. Aristizabal A. y Alberto Galvis C. [14], en la Universidad Nacional de Ingeniería (Lima – Perú) sobre “Análisis de Sistema de Costos Económicos en los Astilleros Menores y Propuestas de Nuevas Innovaciones Lima – Callao” por Víctor Eleno Obregón Pizarro [35], y en la Universidad del Zulia (Maracaibo, Venezuela) sobre “Determinación y análisis de costos por el funcionamiento, operación y mantenimiento del sistema piloto de lagunas de estabilización de la Universidad del Zulia” por Alberto Trujillo y otros investigadores. [36]

En Cuba también se han desarrollado sistemas de este tipo en Sancti – Spíritus sobre “El Análisis Económico Financiero como Herramienta de Dirección. Generalidades” por Yenisbel Cabrera Jiménez y otros investigadores [37], en Las Tunas sobre “Análisis Económico Financiero en la Empresa de Materiales de Construcción de Las Tunas” por Eliades Corazón Del Sol [22], y sobre “Análisis Económico Financiero en la Sucursal Cimex de Las Tunas” por Migdalia Álvarez Peña, Saskia Cosano Delgado y Yamilet Chacón Pupo. [21]

Toda empresa tiene sus particularidades, no se pueden hacer generalizaciones de los estudios que se han realizado, solo por el simple hecho de encontrarse en el mismo lugar. Cada una tiene ciertas características que son necesarias analizar y debido a que en Cuba los estudios que se han realizado sobre la teoría de los conjuntos aproximados han sido limitados, constituyendo generalmente estudios teóricos, se hace necesario aplicar esta teoría en las investigaciones de carácter productivo donde tiene un amplio campo de aplicación.

La Universidad de las Ciencias Informáticas (UCI) constituye un centro de nuevo tipo, donde la producción marcha paralelamente con la docencia. Esto hace que la misma se comporte en gran medida como una empresa productora de Software, donde cada día se incrementan más sus proyectos productivos e investigativos, existiendo ofertas y demandas vinculadas a la producción.

En la actualidad el análisis de costo de estos proyectos se lleva a cabo mediante la técnica tradicional de contabilidad y finanzas. Los gastos de los proyectos se controlan a nivel central, de esta manera los

mismos son rentables. Al realizar este control de forma global y no en cada proyecto se dificulta el análisis local llevando implícito la no confiabilidad de la rentabilidad, deseándose conocer esta información.

Problema científico

¿Cómo desarrollar la propuesta del análisis de costo del proyecto SCADA de la facultad 5?

Objeto de estudio

Aplicación de la teoría de los conjuntos aproximados en la selección de rasgos comunes de un sistema de información.

Objetivo general

Desarrollar una aplicación que facilite el análisis de costo del proyecto SCADA mediante la teoría de los conjuntos aproximados.

Campo de acción

Aplicación de la teoría de los conjuntos aproximados en el análisis de costo del proyecto productivo SCADA de la facultad 5.

Tareas investigativas

- Revisar la bibliografía relacionada con la teoría de los conjuntos aproximados para el análisis del estado del arte.
- Investigar sobre la teoría de los conjuntos aproximados para su aplicación en la investigación.
- Realizar entrevistas a especialistas en teoría de conjuntos aproximados y contabilidad – finanzas para recopilar información relacionada con el tema.
- Aplicar elementos del método Delphi de la consulta de expertos para validar la importancia y necesidad del tema de investigación.
- Realizar el análisis estadístico aplicando el software Statgraphics.

- Seleccionar y utilizar el lenguaje de programación para la implementación de la aplicación.
- Obtener las reglas de producción utilizando el software Weka.
- Desarrollar una aplicación, empleando la teoría de los conjuntos aproximados para facilitar la realización del análisis de costo del proyecto productivo SCADA.

Idea a Defender

Si se logra confeccionar una aplicación que permita el análisis de costo mediante la teoría de conjuntos aproximados, entonces se podrá contribuir al análisis de costo del proyecto productivo SCADA.

Métodos de Investigación

Métodos teóricos:

Analítico – Sintético: Se utilizó para concretar y resumir el conocimiento reflejado en los materiales consultados sobre el tema de la teoría de los conjuntos aproximados y análisis de costo con el objetivo de integrar esa información e ir construyendo el estado del arte y los otros aspectos básicos de esta investigación.

Histórico – Lógico: Se utilizó para estudiar la evolución y desarrollo de las técnicas de Inteligencia Artificial (específicamente de la teoría de los conjuntos aproximados) y la contabilidad con el fin de conocer su estado actual.

Inductivo - Deductivo: Se utilizó para obtener reglas de datos iniciales para llegar a conclusiones.

Métodos empíricos:

Encuesta: Se utilizó en la realización de entrevistas a expertos con el fin de obtener información importante para los objetivos de la investigación.

Experimento: Se utilizó para obtener la efectividad de la teoría de conjuntos aproximados en el análisis de costo de los proyectos productivos.

El documento consta de un Resumen, Introducción, tres Capítulos, Conclusiones Generales, Recomendaciones, Anexos, Citas bibliográficas, Bibliografía consultada y Glosario de términos.

Capítulo 1 Fundamentación Teórica: se realiza un estudio de las técnicas de Inteligencia Artificial y análisis de costo, en el que se tienen en cuenta sus definiciones y aplicaciones.

Capítulo 2 Aplicación del conocimiento: se valida la importancia y necesidad del tema de investigación mediante la consulta de expertos, en particular el método Delphi. Se elabora una Base de Hechos a partir de varias plantillas oficiales empleadas en la UCI y de los gastos en los que incurre el proyecto productivo SCADA, además se dan los elementos fundamentales para el desarrollo de la solución propuesta vinculando la fundamentación teórica con la propuesta de solución.

Capítulo 3 Propuesta de Solución: se trabaja en la búsqueda de la solución del problema, se elabora una aplicación a partir de un conjunto de reglas de producción obtenidas en el software WEKA, además de describir las funcionalidades y las pruebas que se le realizaron al software, así como los resultados obtenidos.

CAPÍTULO I FUNDAMENTACIÓN TEÓRICA

1.1 Introducción

En este capítulo se tratan aspectos y conceptos importantes que ayudan a comprender el tema a tratar en la investigación. Se centrará en los principales conceptos y técnicas mediante las cuales sean posibles las decisiones financieras, así como las principales características de los conjuntos aproximados. Además, se tratarán algunas generalidades de la Inteligencia Artificial.

1.2 Análisis de costo.

El análisis de los proyectos constituye la técnica matemático-financiera y analítica, a través de la cual se determinan los beneficios o pérdidas en los que se puede incurrir al pretender realizar una inversión o algún otro movimiento, en donde uno de sus objetivos es obtener resultados que apoyen la toma de decisiones referente a actividades de inversión.

Asimismo, al analizar los proyectos de inversión se determinan los costos de oportunidad en que se incurre al invertir, para obtener beneficios al instante, mientras se sacrifican las posibilidades de beneficios futuros, o si es posible privar el beneficio actual para trasladarlo al futuro, al tener como base específica a las inversiones.

Una de las evaluaciones que deben de realizarse para apoyar la toma de decisiones en lo que respecta a la inversión de un proyecto, es la que se refiere a la evaluación financiera, que se apoya en el cálculo de los aspectos financieros del proyecto.

En las circunstancias por las que atraviesa el país, la economía está orientada hacia su total recuperación, para ello la elevación de la eficiencia en la productividad social se deberá convertir en la vía principal para asegurar el desarrollo económico. Constituye de gran importancia conocer el comportamiento de la eficiencia económica financiera con vista a tomar decisiones adecuadas inherentes al uso de los recursos materiales y financieros.

Miguel Ángel Benítez en 1997, en su libro Contabilidad y Finanzas para la Formación Económica de los Cuadros de Dirección, expresa que: “La medición de la eficiencia económica operativa con que se explota una entidad se realiza básicamente mediante el análisis financiero y económico operativo de sus

actividades". [6]

El correcto funcionamiento de cualquier sistema lo constituye el análisis de estados financieros, también conocido como análisis económico – financiero, análisis de balance o análisis contable, que es un conjunto de técnicas utilizadas para diagnosticar la situación y perspectivas de la empresa con el fin de poder tomar decisiones adecuadas, estudia la estructura y evolución de los resultados de la empresa (ingresos y gastos) y de la rentabilidad de los capitales utilizados. [21]

1.3 Conceptos relacionados con Análisis de costo.

En este epígrafe se dará una breve explicación de los conceptos que se relacionan con Análisis de costo, por ejemplo: sistemas de costos, solvencia, rentabilidad del producto y la toma de decisión, reflejando cómo están relacionados estos conceptos.

1.3.1 Costo

Es la suma de los esfuerzos y recursos que se han invertido para producir algo. [3]

Es lo que hay que entregar para conseguir algo, lo que es preciso pagar o sacrificar para obtenerlo, ya sea mediante la compra, el intercambio o la producción. En este último caso el costo representa lo que hay que entregar a cambio para obtener los diversos insumos que se necesitan para su producción. [12]

El cálculo de los costos puede mejorar la toma de decisiones, contribuye a reducir el costo y el tiempo de ejecución además de mejorar la calidad y el rendimiento del producto final. Permite el análisis de los gastos incurridos en la producción de un producto o varios, así como de componentes de éstos, de la producción en proceso e incluso de la producción defectuosa.

La finalidad del cálculo del costo es:

- Garantizar el análisis de la eficiencia de la producción, facilitando la adopción de medidas oportunas para corregir las desviaciones del costo planificado e incrementar la eficiencia de las empresas.
- Contribuir a la correcta fundamentación de los indicadores del plan.
- Facilitar el análisis en la formación de los precios.

La gestión de costos está integrada por tres procesos:

1. Estimación de costos: Es el proceso donde se desarrolla una aproximación de los costos de los recursos necesarios para las actividades del proyecto.
2. Preparación del presupuesto de costos: Es el proceso donde se determinan los costos estimados de actividades individuales o paquetes de trabajo a fin de establecer una línea base de costo.
3. Control de costos: Proceso que permite monitorear las variaciones del costo y controlar los cambios en el presupuesto del proyecto.

1.3.2 Sistemas de costos.

El sistema de costo es un conjunto de métodos, normas y procedimientos que rigen la planificación, determinación y análisis del costo, así como el proceso del registro de los gastos de una o varias actividades productivas en una empresa, de forma interrelacionada con los subsistemas que garantizan el control de la producción y de los recursos materiales, laborales y financieros. [15]

Este sistema debe garantizar los requerimientos informativos para una correcta dirección de la empresa, así como para los niveles intermedios y superiores de dirección.

El sistema debe ante todo ser flexible y dinámico en cuanto a permitir a las empresas la introducción de las exigencias informativas y de análisis que asegure una correcta toma de decisiones, por lo cual deben ser ellas quienes lo confeccionen y lo adecúen a sus necesidades, tomando en cuenta los aspectos organizativos y procesos tecnológicos específicos de cada una.

1.3.3 Presupuesto de costo.

En toda actividad productiva al fabricar un producto o prestar un servicio se generan costos, entendiéndose que los costos son desembolsos monetarios relacionados justamente con la fabricación del producto o la prestación del servicio ya sea en forma directa o indirectamente. [20]

1.3.3.1 Importancia de los Costos.

- Proporciona información concerniente a las fases de planeación de los proyectos.
- El mejor control de las operaciones internas de la empresa.

- La toma de decisiones.
- La formación correcta de los estados financieros.
- La elección de alternativas. [3]

1.3.3.2 La estimación del costo de un proyecto.

La estimación de costo de un proyecto consiste en estimar los costos de los recursos necesarios (humanos y materiales) para completar las actividades del proyecto. En la aproximación de costos la persona que estima considera las posibles variaciones del estimado final con propósito de mejorar la administración del presupuesto del proyecto. [24]

Cuando un proyecto se realiza bajo contrato se debe tener cuidado en distinguir el costo estimado del precio:

1. Costo estimado: ¿cuánto le costará a la organización que realiza el proyecto proveer el producto o servicio? El costo estimado es un cálculo económico.
2. Precio: ¿cuánto recargará la organización que realiza el proyecto por el producto o servicio? El precio es una decisión de negocios.

La estimación de costos incluye la identificación y consideración de varias alternativas de costo, y esto es una decisión gerencial. Por ejemplo, realizar trabajo adicional durante la fase de diseño debido a que esto tiene el potencial de reducir el costo en la fase de ejecución. [24]

1.3.3.3 Costos de producción.

Los costos de producción (también llamados costos de operación) son los gastos necesarios para mantener un proyecto, línea de procesamiento o un equipo en funcionamiento. [4] Expresan la magnitud de los recursos materiales, laborales y monetarios necesarios para alcanzar un cierto volumen de producción con una determinada calidad.

Representa todas las operaciones realizadas desde la adquisición de la materia prima, hasta su transformación en artículo de consumo o de servicio, integrado por:

- Materia Prima: Es el elemento que se convierte en un artículo de consumo o de servicio.
- Sueldos y Salarios: Es el esfuerzo humano necesario para la transformación de la materia prima.

- **Gastos Indirectos de Producción:** Son los elementos necesarios para la transformación de la materia prima, como son: el lugar en donde se trabaja, el equipo, las herramientas, etc.

El costo de producción tiene dos características opuestas, que algunas veces no están bien entendidas en los países en vías de desarrollo. La primera es que para producir bienes uno debe gastar; esto significa generar un costo. La segunda característica es que los costos deberían ser mantenidos tan bajos como sean posibles y eliminar los innecesarios.

1.3.3.3.1 Clasificación de los costos de producción.

Los costos de producción pueden dividirse en dos grandes categorías: **costos directos o variables**, que son proporcionales a la producción, como materia prima, y los **costos indirectos**, también llamados **fijos** que son independientes de la producción, como los impuestos que paga el edificio. Algunos costos no son ni fijos ni directamente proporcionales a la producción y se conocen a veces como semivariable.

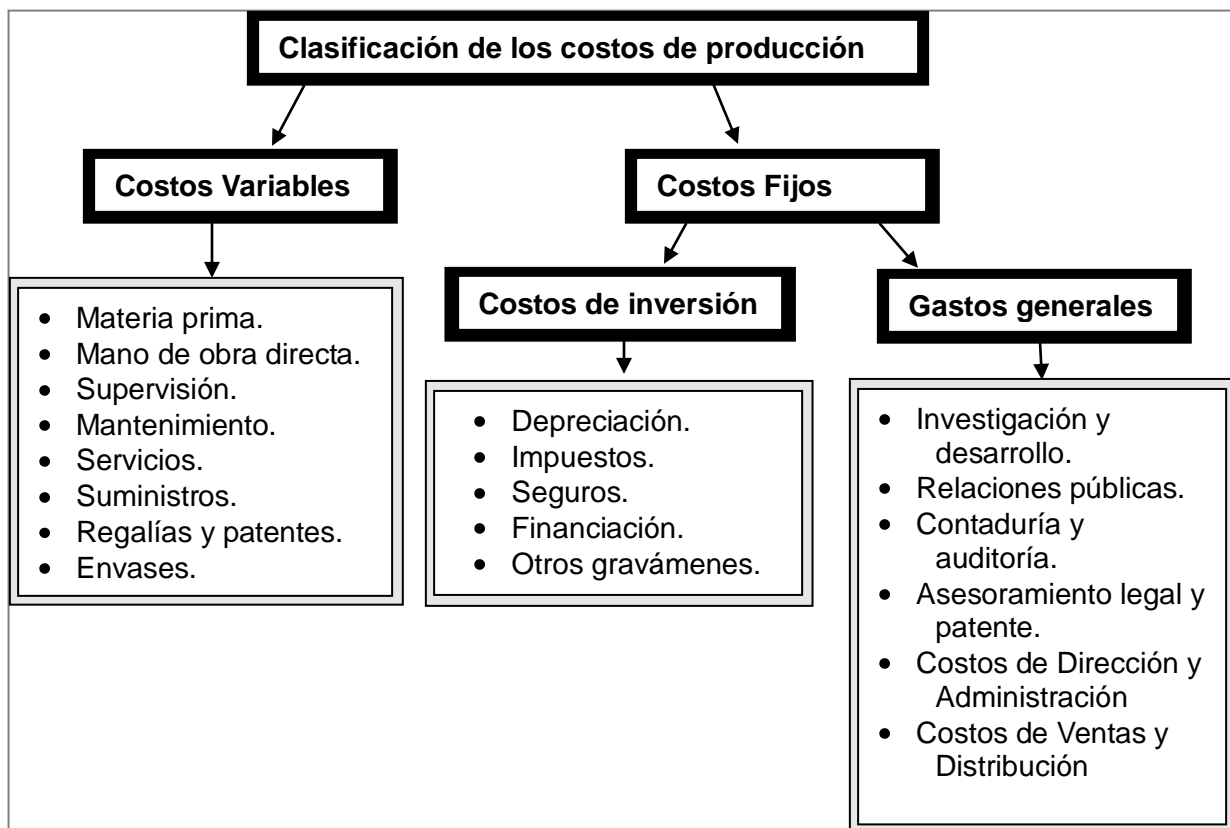


Figura 1: Clasificación de los costos de producción.

Fuente: Elaboración propia.

1.3.4 Solvencia.

Es la capacidad financiera (capacidad de pago) de la empresa para cumplir sus obligaciones de vencimiento a corto plazo y los recursos con que cuenta para hacer frente a tales obligaciones, o sea una relación entre lo que una empresa tiene y lo que debe. [22]

Para que una empresa cuente con solvencia, debe estar capacitada para liquidar los pasivos contraídos, al vencimiento de los mismos y demostrar también mediante el estudio correspondiente que podrá seguir una trayectoria normal que le permita conservar dicha situación en el futuro.

Se debe diferenciar el concepto de solvencia con el de liquidez, debido a que liquidez es tener el efectivo necesario en el momento oportuno que permita hacer el pago de los compromisos anteriormente contraídos.

Solvencia es contar con los bienes y recursos suficientes para respaldar los adeudos que se tengan contraídos, aún cuando estos bienes sean diferentes al efectivo.

Es decir, liquidez es cumplir con los compromisos y solvencia es tener con que pagar esos compromisos, para tener liquidez se necesita tener solvencia previamente. [26]

También lo anterior conduce a que la solvencia es la posesión de abundantes bienes para liquidar deudas pero si no hay facilidad para convertir esos bienes en dinero o efectivo para hacer los pagos entonces no existe la liquidez.

Activo: Representa todos los bienes y derechos que son propiedad de la empresa.

Pasivo: Representa todas las deudas y obligaciones a cargo de la empresa.

1.3.5 Precio de venta.

Es el valor de los productos o servicios que se venden a los clientes. [27]

Los principales factores que deben ser considerados antes de establecer el precio son:

- Costos de producción.
- Precios de la competencia.
- Demanda del producto y comportamiento del consumidor.

- Imagen deseada por la empresa.
- Capacidad adquisitiva del consumidor.
- Limitaciones legales y gubernamentales.

El precio de venta se efectúa en dependencia de la empresa, atendiendo a los factores anteriores, según los que se tengan en cuenta será la manera de realizar su cálculo.

1.3.6 Utilidad o pérdida.

Es el excedente que se obtiene luego de cubrir los costos correspondientes para la elaboración de un producto o dar un servicio.

La Utilidad Neta se obtiene de restar al Precio de venta del producto su coste unitario.

Utilidad Neta = Precio de venta – Costo de producción. [23]

Si el resultado del cálculo da como resultado un valor positivo es porque existen ganancias, de no ser así entonces existirían pérdidas.

1.3.7 Margen de utilidad.

Es la diferencia entre ingresos y gastos totales, incluyendo impuestos, expresadas en un porcentaje de los ingresos.

El Margen neto de utilidad se obtiene de dividir el Utilidad Neta entre Precio de ventas.

Margen neto de utilidad = Utilidad Neta/Precio de ventas. [26]

Se muestra por cada unidad monetaria de las Ventas cuanto se convierte en Utilidad Neta.

1.3.8 Toma de decisión.

Como uno de los objetivos del análisis de costo es obtener resultados que apoyen la toma de decisiones referente a actividades de inversión es importante tener conocimiento sobre la misma.

Es el conjunto de acciones adoptadas en un momento específico, como resultado de la aplicación de ciertas reglas y políticas a las condiciones particulares existentes en dicho momento. [25]

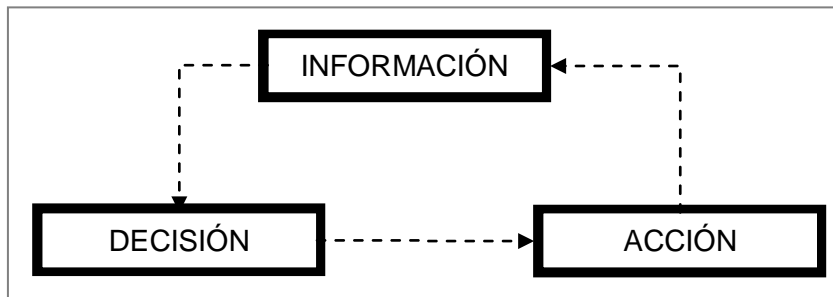


Figura 2: Toma de decisiones.

Fuente: Elaboración propia.

Etapas de la toma de decisiones:

- a) Definición del problema.
- b) Análisis de la información disponible.
- c) Desarrollo de las soluciones alternativas.
- d) Selección de la decisión.
- e) Implantación de la estrategia elegida.

1.3.9 Rentabilidad del producto.

La rentabilidad mide la eficiencia con la cual una empresa utiliza sus recursos financieros. Una empresa es eficiente cuando no desperdicia recursos. Cada empresa utiliza recursos financieros para obtener beneficios. Estos recursos son, por un lado, el capital (que aportan los accionistas) y, por otro, la deuda (que aportan los acreedores). A esto hay que añadir las reservas: los beneficios que ha retenido la empresa en ejercicios anteriores con el fin de autofinanciarse (estas reservas, junto con el capital, constituyen los “Fondos Propios”).

En realidad, hay varias medidas posibles de rentabilidad, pero todas tienen la siguiente forma:
 Rentabilidad = Beneficio / Recursos Financieros. [13]

Un software va a ser rentable cuando por cada peso que se invierte se obtiene un margen de utilidad.

1.4 Inteligencia Artificial.

En el campo de las finanzas existe una fuerte motivación orientado a la construcción de sistemas que

incorporen conocimiento, permitiendo tomar de decisiones eficientes y oportunas en el análisis de costo; por lo que no se puede dejar de mencionar las técnicas de IA aplicadas al aprendizaje automático y gestión del conocimiento. La Inteligencia Artificial trata de conseguir que los ordenadores simulen en cierta manera la inteligencia humana. Se acude a sus técnicas cuando es necesario incorporar en un sistema informático, conocimiento o características propias del ser humano.

A continuación se describen algunos conceptos de Inteligencia Artificial desde el punto de vista de diferentes autores: [29]

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

“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.

“IA... está relacionada con conductas inteligentes en artefactos”. Nilsson, 1998.

1.4.1 Algunos conceptos relacionados con Inteligencia Artificial.

En este epígrafe se dará una breve explicación de algunos de los conceptos que se relacionan con Inteligencia Artificial, por ejemplo: aprendizaje automático, árbol de decisión, minería de datos y selección de rasgos dando a conocer como están relacionados.

1.4.1.1 Aprendizaje automático.

El aprendizaje automático es una rama de la inteligencia artificial cuyo objetivo es que las computadoras sean capaces de aprender. En realidad, se trata de que los programas de ordenador sean capaces de inducir conocimiento de cierta información, generalmente conjuntos de ejemplos, que utilizarán posteriormente a su favor. Por ejemplo, en un juego por computador un agente aprende si es capaz de ajustar su estrategia para adaptarse a diferentes oponentes.

Visto así, el aprendizaje automático puede considerarse compuesto por dos fases: una primera en la que elige/observa las características más relevantes de un evento y otra, en la que compara estas características y las adapta a su modelo.

El algoritmo de aprendizaje automático, independientemente del tipo de aprendizaje que lleve a cabo, acepta una información como entrada y genera un modelo que intenta explicar los datos de entrada; pudiendo estar representado dicho modelo, por ejemplo, mediante conjuntos de reglas. [32]

1.4.1.1.1 Algoritmos de aprendizaje automático.

Existe una gran cantidad de algoritmos de aprendizaje automático que, en función de los modelos que proporcionan, son clasificados en algoritmos de clasificación, agrupamiento (clustering) y asociación. Los algoritmos de clasificación proporcionan modelos que clasifican las nuevas instancias con un valor concreto para su clase, nominal o numérica. Pueden distinguirse varios tipos de algoritmos de clasificación en función de las características del modelo que proporcionan o de la fundamentación teórica en que se sustentan. A continuación se dará una breve descripción de los utilizados en esta investigación

Algoritmos de reglas: El modelo que proporcionan son una regla o conjunto de reglas ordenadas por importancia que clasifican una determinada instancia cuando sus atributos cumplen una regla.[32]

NNge.

Usa un algoritmo similar al vecino más cercano no anidado de forma que los hiperrectángulos pueden ser vistos como condiciones IF-THEN. [2]

Algoritmos de árboles: Estos algoritmos generan un modelo en árbol (árbol de decisión o diagrama en árbol).

Características básicas de los árboles de decisión.

Un «árbol de decisión» es un diagrama de flujo, con estructura de árbol, en donde los nodos internos representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones, y los “nodos hoja” representan las clases. El nodo en la parte superior del árbol se le conoce como nodo raíz.

[34]

Para clasificar una instancia “desconocida”, se sigue el flujo del árbol hacia abajo, de acuerdo a los valores que tengan los atributos para cada nodo, y cuando se llega a un “nodo hoja”, la instancia se clasifica de acuerdo a la clase asignada por dicho nodo.

Existen diversos métodos para la inducción de árboles de decisión (ID3, C4, C4.5, Bayesiano, CART, etc.), cada uno de ellos ofrece diferentes capacidades, pero en general, dichos algoritmos son apropiados para solucionar «problemas de clasificación». Estos problemas presentan las siguientes características:

- El objeto de estudio del problema está representado por un conjunto fijo de atributos.
- La función objetivo del problema tiene valores discretos.
- Los atributos son descritos mediante conceptos disyuntivos.
- Puede haber errores en los datos de entrenamiento, tanto en la clasificación como en los valores de los atributos.
- Los datos de entrenamiento, con los cuales se construye el modelo, pueden tener valores desconocidos para algunos atributos.

Es importante señalar que los árboles de decisión presentan una gran ventaja respecto a otras técnicas de clasificación. Esta ventaja consiste en poder representar al conocimiento obtenido mediante el uso de reglas de decisión. [34]

ADTree - Alternating Decision Tree.

Un Árbol de decisión alternativo es un método de clasificación proveniente del aprendizaje automático conocido en inglés como Alternating Decision Tree (ADTree). Las estructuras de datos y el algoritmo son una generalización de los árboles de decisión. [33]

Los ADTree contienen nodos splitter y nodos de predicción. El primero es un nodo que es asociado con una prueba, mientras que un nodo de predicción es asociado con una regla. Una instancia define una serie de caminos en un ADTree. La clasificación es asociada con una instancia que es el signo de la suma de las predicciones cercanas al camino en el que es definido por esta instancia. El ADTree puede ser visto como una consistencia de una raíz nodo de dirección y cuatro unidades de tres nodos cada uno. Cada

unidad es una regla de decisión y está compuesta por un nodo splitter y dos nodos de predicción que son sus hijos. [33]

Las reglas en un ADTree son similares a las de árboles de decisión, consecuentemente se puede aplicar métodos de empuje o aumento en el orden para diseñar un algoritmo de aprendizaje ADTree. Básicamente, el algoritmo de aprendizaje para construcción de un ADTree es una estrategia top-down. Cada paso de aumento es seleccionado y adiciona una nueva regla o su equivalente a una nueva unidad consistente de un nodo splitter y dos nodos de predicción. [33]

ID3.

Id3 particiona el espacio de manera lineal teniendo como objetivo maximizar la ganancia de información (reducir la entropía) con cada nueva división. [32]

El ID3 es un algoritmo simple pero potente, cuya misión es la elaboración de un árbol de decisión bajo las siguientes premisas:

- Cada nodo corresponde a un atributo y cada rama al valor posible de ese atributo. Una hoja del árbol especifica el valor esperado de la decisión de acuerdo con los ejemplos dados. La explicación de una determinada decisión viene dada por la trayectoria desde la raíz a la hoja representativa de esa decisión.
- A cada nodo es asociado aquel atributo más informativo que aún no haya sido considerado en la trayectoria desde la raíz.
- Para medir cuánto de informativo es un atributo se emplea el concepto de entropía. Cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será el atributo para la clasificación. [33]

El ID3 es capaz de tratar con atributos cuyos valores sean discretos o continuos. En el primer caso, el árbol de decisión generado tendrá tantas ramas como valores posibles tome el atributo.

El aprendizaje automático puede ser utilizado cuando concurren alguna o varias de las siguientes características:

- La aplicación es difícil de desarrollar (reconocer complejos patrones: caras).
- Aplicaciones autoadaptables, tales como asistentes personales, filtros anti-spam.

- Minería de datos, es decir, se desea llevar a cabo un análisis profundo de un conjunto de datos. [32]

1.4.1.2 Minería de datos.

La minería de datos (*Data Mining*, en inglés) engloba un gran conjunto de técnicas y métodos que tienen como objetivo extraer información no trivial que se cumple en un conjunto de datos. Es decir, convertir los datos de entrada en conocimiento para permitir la toma de decisiones. La fundamentación teórica de la minería de datos se encuentra en la inteligencia artificial, principalmente el aprendizaje automático, las bases de datos y la estadística. [32]

1.4.1.3 Selección de rasgos

Representa el problema de encontrar un subconjunto óptimo de características (rasgos o atributos) en una base de datos y según cierto criterio, tales que se pueda generar un clasificador con la mayor calidad posible a través de un algoritmo inductivo que corra sobre los datos, pero solo tomando en cuenta el subconjunto obtenido.

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 encargado de generar los subconjuntos de rasgos; dados N rasgos se tiene $2^N - 1$ posibles subconjuntos. Seleccionar los rasgos relevantes de un conjunto de datos es una tarea necesaria en el aprendizaje automatizado, dada su importancia en el descubrimiento de reglas y relaciones en grandes volúmenes de datos entre otras aplicaciones.

Una herramienta matemática muy potente para la selección es la teoría de conjuntos aproximados.

El concepto de atributo será tratado en el epígrafe siguiente. [38]

1.4.2 Teoría de los conjuntos aproximados.

La teoría de los conjuntos aproximados (*Theory Rough Set*, en inglés) ha tenido un gran desarrollo y aplicación en diferentes campos del conocimiento. Esta teoría fue desarrollada a mediados de los años ochenta por el científico Polaco Zdzisaw Pawlak para analizar datos donde existe alta ambigüedad, permitiendo su clasificación y toma de decisiones. Su propósito principal es la transformación de datos en conocimiento, es decir, se trata de extraer información útil de las bases de datos. Esta teoría resulta muy útil para descubrir dependencias entre los atributos de las observaciones, evaluar el nivel de significación

y para el tratamiento de datos o información inconsistente. [31]

Existen 3 componentes principales de estudio de la IA que son: el conocimiento, los algoritmos heurísticos y la incertidumbre. Luego aparece otro campo denominado Aprendizaje Automático y más tarde la Teoría de los Conjuntos Aproximados (Rough Sets Theory) RST, la cual crea una mezcla entre el aprendizaje automático y la incertidumbre. Figura 3.

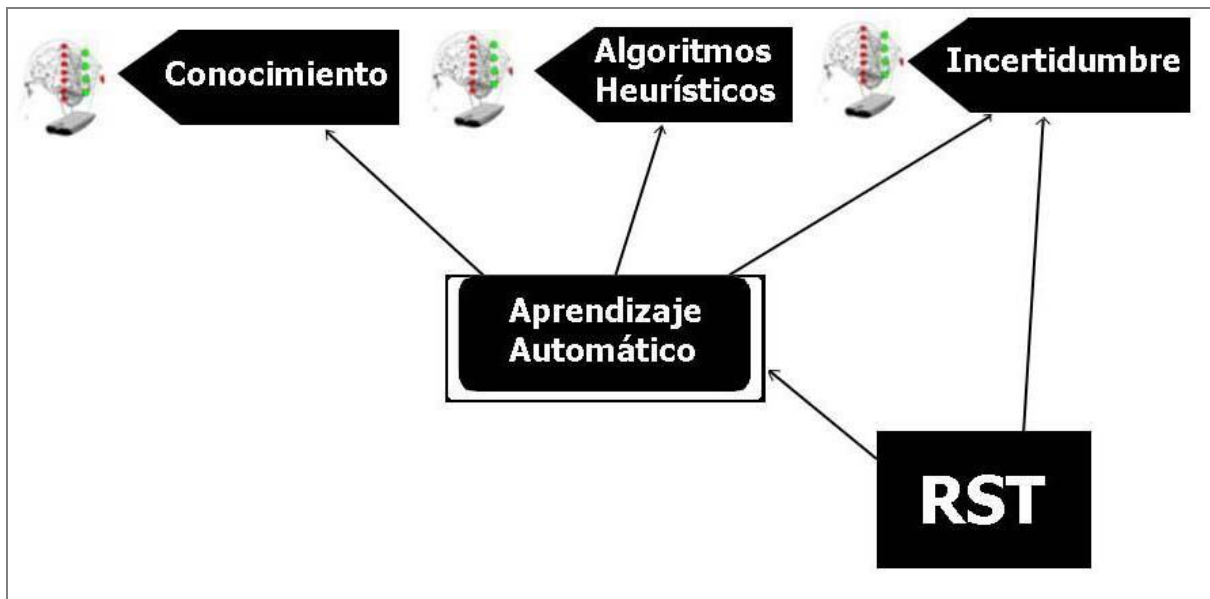


Figura 3: La teoría de conjuntos aproximados en el contexto de la Inteligencia Artificial.

La teoría de los Rough Sets (RST) asume la representación del conocimiento en forma de tabla de decisión, que es una forma particular de un sistema de información.

Representación de la RST:

RST= *Sistema de información + Relación de separabilidad.*

Dónde *Sistema de información* no es más que los datos del dominio, y la *Relación de separabilidad* está constituida por las componentes de la teoría.

1.4.2.1 Sistema de Información.

En la tabla de información se representa el conocimiento de los objetos. En las filas de la tabla se indican los objetos (acciones, alternativas, candidatos, pacientes, empresas, etc.), mientras que las columnas se

corresponden con los atributos. Las entradas en la tabla son los valores del atributo. Es decir, la entrada en columna q y en fila x tiene el valor f(x, q). Por tanto, para cada par (objeto, atributo) se conoce un valor denominado descriptor. Cada fila de la tabla contiene descriptores que representan información correspondiente a un objeto. La relación de no-diferenciación ocurriría si dados dos objetos para todos los atributos tomaran los descriptores el mismo valor.

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

Tabla 1: Representación de un sistema de información.

1.4.2.2 Sistema de decisión.

Si a cada elemento de U se le agrega un nuevo atributo "d" llamado decisión indicando la decisión tomada en ese estado o situación entonces se obtiene un Sistema de decisión (U, A ∪ {d}, donde d ∉ A). [30]

Consistencia del sistema de decisión:

Cuando dos objetos inseparables, según un subconjunto de atributos B, pertenecen a clases de decisión diferentes se dice que el Sistema de decisión es inconsistente, en otro caso es consistente. También se conocen como sistemas no deterministas o deterministas, respectivamente.

Se define entonces la función $\delta_B(x)$, la cual da la relación de clases de decisión a las cuales pertenecen los objetos inseparables del objeto x según B ($I_B(x)$): [30]

$$\delta_B(x) = \{ v \in V_d : \exists y \in I_B(x) \text{ tales que } d(y) = v \}$$

Un sistema de decisión es consistente, si y sólo si, el conjunto $\delta_B(x)$ es un conjunto unario ¹ para todo $x \in U$.

$$|\delta_B(X)| > 1 \Rightarrow \text{S.I. NO determinista}$$

Los sistemas de decisión pueden construirse con más de un atributo de decisión, es decir, se tiene el conjunto de atributos de decisión D. En este caso el sistema se define por (U, A ∪ D).

¹ Conjunto unitario: es el conjunto cuyo único elemento es x, lo notamos {x}.

Al igual que se ha hablado de relaciones de separabilidad, se puede hablar de relación de inseparabilidad; similarmente al caso en que hablamos de funciones de semejanzas o funciones de distancias para medir la cercanía entre dos objetos.

Relación de inseparabilidad.

A cada subconjunto de atributos B de A ($B \subseteq A$) está asociada una relación binaria de inseparabilidad denotada por I_B , la cual es el conjunto de pares de objetos que son inseparables uno de otro por esa relación. [30]

$$I_B = \{ (x,y) \in U \times U: f(x,A_i) = f(y,A_i) \text{ para todo } A_i \in B \}$$

Si $(x,y) \in I_B$ se dice que los objetos x e y son B - inseparables.

Una relación de inseparabilidad (indiscernibility relation, en inglés) que sea definida a partir de formar subconjuntos de elementos de U que tienen igual valor para un subconjunto de atributos B de A ($B \subseteq A$) es una relación de equivalencia. Es decir, es una relación binaria $R \subseteq U \times U$ que es reflexiva, simétrica y transitiva.

La complejidad computacional en tiempo de construir una relación de equivalencia es $O(lm^2)$, donde l es el número de atributos que describen los objetos y m es la cantidad de objetos en el universo.

La clase de equivalencia de un elemento x de U es el conjunto de todos los elementos y de U tal que xRy ; es decir, los elementos de U que son similares a x considerando los atributos contenidos en B . Una relación de equivalencia induce una partición de el universo U . Por $B(x)$ se denota al conjunto de todos los $y \in U$ tal que yRx ; es decir, el elemento de la partición al que pertenece el objeto x .

1.4.2.3 Aproximaciones de un conjunto.

Sea el sistema de información (U,A) , y los conjuntos $B \subseteq A$ y $X \subseteq U$. Se puede aproximar X usando solamente la información contenida en B construyendo dos conjuntos llamados aproximación inferior (B_*) y aproximación superior (B^*) respectivamente del conjunto X para la relación B . Un conjunto aproximado es cualquier subconjunto $X \subseteq U$ definido a través de sus aproximaciones inferior y superior.

La aproximación inferior de un conjunto (con respecto a un conjunto dado de atributos) se define como la

colección de casos cuyas clases de equivalencia están contenidas completamente en el conjunto; mientras que la aproximación superior se define como la colección de casos cuyas clases de equivalencia están al menos parcialmente contenidas en el conjunto. Formalmente:

$$B_*(X) = \{x \in U \mid B(x) \subseteq X\}$$

$$B^*(X) = \{x \in U \mid B(x) \cap X \neq \emptyset\}$$

Los elementos de $B_*(X)$ son todos y solamente aquellos objetos del universo U los cuales pertenecen a las clases de equivalencia generadas por la relación I_B contenidas en X ; mientras que los elementos de $B^*(X)$ son todos y solamente aquellos objetos de U los cuales pertenecen a las clases de equivalencia generadas por la relación de inseparabilidad conteniendo al menos un objeto x perteneciente a X . A partir de ellos se define la región límite de X para la relación de equivalencia B :

Región límite.

$$BN_B(X) = B^*(X) - B_*(X)$$

Si el conjunto BN_B es vacío entonces el conjunto X es exacto respecto a la relación de equivalencia B . En caso contrario, $BN_B(X) \neq \emptyset$, el conjunto X es inexacto o aproximado con respecto a B . El principal objetivo de esta teoría es la síntesis de aproximaciones para los conceptos. [30]

La complejidad computacional de encontrar el conjunto $B^*(X)$ o el conjunto $B_*(X)$ es $O(lm^2)$, donde l es el número de atributos que describen los objetos y m es la cantidad de objetos en el universo.

1.4.2.4 Algunas propiedades de los conjuntos aproximados.

a) $B_*(X) \subseteq X \subseteq B^*(X) \subseteq U$

b) $B_*(X) \cap B_*(Y) = \emptyset$

c) $B_*(X) \cap B^*(Y) = \emptyset$

d) $B_*(\emptyset) = B^*(\emptyset) = \emptyset$

e) $B_*(U) = B^*(U) = U$

f) $B^*(X \cup Y) = B^*(X) \cup B^*(Y)$

g) $B_*(X \cap Y) = B_*(X) \cap B_*(Y)$

h) $X \subseteq Y \Rightarrow B_*(X) \subseteq B_*(Y) \text{ y } B^*(X) \subseteq B^*(Y)$

i) $B_*(X) = U - B^*(U - X)$, propiedad de complementariedad. [30]

1.4.2.5 Medidas asociadas a conjuntos aproximados.

Sobre la base del conocimiento en B los objetos que pertenecen a $B_*(X)$ pueden ser clasificados con certeza como miembros de X, los objetos miembros de $B^*(X)$ pueden ser solamente clasificados como posibles miembros de X. Los elementos en B_N no pueden ser clasificados con absoluta certeza como miembros del conjunto X. Los elementos del conjunto $U - B^*(X)$ con certeza no pertenecen a X. [30]

Por eso, los conjuntos aproximados pueden ser vistos como un modelo matemático de los conceptos vagos, es decir, aquellos en los que los elementos del universo no pueden ser clasificados con certeza como elementos o no del concepto. Las principales medidas construidas a partir de los conjuntos aproximados son las siguientes.

Precisión de la aproximación.

La vaguedad del concepto o conjunto X, también llamada precisión de la aproximación con respecto a la relación B se puede caracterizar matemáticamente por el coeficiente:

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

Donde $|X|$ denota la cardinalidad del conjunto X, y $0 \leq \alpha_B(X) \leq 1$. Si $\alpha_B(X) = 1$, el conjunto X será duro o exacto con respecto a la relación de equivalencia B, mientras que si $\alpha_B(X) < 1$, el conjunto X es aproximado o vago con respecto a B. Esta magnitud mide el grado de perfección o integridad del conocimiento sobre el conjunto X considerando los atributos incluidos en la relación de equivalencia. [30]

Calidad de la aproximación.

La calidad de la aproximación de X por medio de los atributos en B se define por:

$$\gamma_B(X) = \frac{|B_*(X)|}{|X|}$$

Esta medida de calidad representa la frecuencia relativa de los objetos correctamente clasificados por medio de los atributos en B. Además, $0 \leq \alpha_B(X) \leq \gamma_B(X) \leq 1$. [30]

Calidad de la clasificación.

El coeficiente $\gamma_B(Y)$ se denomina calidad de la clasificación. [30]

$$\gamma_B(Y) = \frac{\sum_{i=1}^n |B_i(Y_i)|}{|U|}$$

Empleando el concepto de inseparabilidad se puede definir la función de pertenencia a conjuntos aproximados. Si el concepto o conjunto X es vago sus contornos no están rigurosamente definidos por lo que surge la incertidumbre relacionada con la cuestión de la pertenencia de los elementos al conjunto.

Función de pertenencia aproximada.

La función de pertenencia de un elemento x a un conjunto X según la relación de equivalencia B se define como:

$$\mu_x^B(x) = \frac{|X \cap B(x)|}{|B(x)|}$$

El grado de pertenencia está en el intervalo [0,1]. La función de pertenencia puede ser comprendida como un coeficiente que certidumbre de que un elemento x sea miembro del conjunto X . Esta medida cuantifica el grado de correspondencia relativa entre el conjunto X y la clase de equivalencia a la cual pertenece x . Se puede interpretar análogamente a la probabilidad condicional y entenderse como el grado de certidumbre de que el elemento x pertenezca al conjunto X . [30]

1.4.2.6 Reglas de decisión.

Una regla de decisión puede expresarse como una implicación lógica de la forma $P \rightarrow Q$ que relaciona los atributos de condición con el de decisión.

El conjunto de reglas generadas puede considerarse como una representación del conocimiento adquirido sobre todos los objetos contenidos en la base de datos. Esta representación se emplea para clasificar nuevos individuos a partir de sus atributos.

Una regla de decisión puede expresarse como una sentencia lógica que relaciona la descripción de condiciones y las clases de decisión. Toma la siguiente forma:

SI <se cumplen condiciones> ENTONCES <el objeto pertenece a una clase de decisión dada>

La teoría de los Rough Sets conforma un esquema formal de las reglas generadas; pueden ser determinísticas o no determinísticas. Son procedentes de una representación del conocimiento en forma de tabla de decisión, dando lugar a la configuración un nuevo paradigma de representación simbólica del conocimiento con derivación inductiva de reglas: aprendizaje inductivo. Por determinística (consistente, precisa, exacta) entendemos si $C \rightarrow D$ y en otro caso, es no determinística (inconsistente, aproximada) que sería cuando las condiciones pueden conducir a varias posibles decisiones. [9] Cada regla de decisión se caracteriza por su fuerza, esto es, el número de objetos que satisfacen la parte de la condición de la regla (en el argot se dice cubiertos por la regla) y que pertenecen a la clase de decisión sugerida por ella.

Según Pawlak y Solowinski (1994), existen diferentes procedimientos que permiten derivar reglas de decisión, sin embargo, los algoritmos existentes usan algunas de las siguientes estrategias:

- Generación de un conjunto mínimo de reglas, cubriendo todos los objetos de una tabla de decisión.
- Generación de un conjunto exhaustivo de reglas, buscando obtener todas las posibles reglas de una tabla de decisión.
- Generación de un conjunto “robusto” de reglas, que eventualmente podría ser discriminante, pero cubriendo a muchos objetos de la tabla de decisión, es decir, no se llega a cubrir necesariamente a todos los objetos.

1.4.2.7 ¿Cómo podemos usar RST para tomar decisiones?

Para tomar decisiones basadas en ejemplos.

Entonces se puede:

- ⇒ Analizar la consistencia del sistema de decisión.
- ⇒ Calcular la *precisión* de cada decisión.
- ⇒ O calcular la *calidad de la clasificación* que se logra con cada decisión.
- ⇒ Medir en qué medida el estado actual donde se toma la decisión se corresponde con “la situación patrón” de cada decisión.

⇓

Con todos estos datos desarrollar un modelo para tomar la decisión correcta.

1.4.2.8 Ventajas de esta teoría en comparación a otras técnicas de la IA.

La teoría de *rough sets* ofrece, fundamentalmente, cierta ventaja en comparación con las redes

neuronales, fuzzy logic y algoritmos genéticos debido a que no se requiere conocimiento previo o dominio de un experto para generar las reglas que la rigen, sino que se pueden usar algoritmos conocidos. Adicionalmente, las reglas que se generan al aplicar esta técnica se traducen en lógica proposicional, las cuales pueden ser evaluadas por un experto. [8]

Otras ventajas que se obtienen con la aplicación de este método son:

- Es útil para analizar sistemas de información que representan el conocimiento adquirido por la experiencia.
- Utilización de información original por lo cual no se necesita de información externa.
- Es un método rápido para procesar gran cantidad de información tanto cuantitativa como cualitativa.
- El conjunto de reglas de decisión derivadas de la aplicación de este método, da una idea generalizada del conocimiento contenido en la base de dato de la cual se extrae la información. Lo anterior elimina la redundancia típica de la información original.
- Elimina las variables redundantes reduciendo el coste, en tiempo y dinero, del proceso de decisión.
- Se obtienen unas reglas de decisión fácilmente comprensibles que no necesitan interpretación de ningún experto.
- Las reglas de decisión se basan en hechos pues están soportadas por un conjunto real de ejemplos y los resultados son fáciles de obtener.
- Las reglas están bien justificadas por extraerse de ejemplos reales lo que justificaría las decisiones que en base a ellas se tome.

Como todas las técnicas de la Inteligencia Artificial tienen sus ventajas y desventajas para darle mayor consistencia y precisión, los investigadores en este campo están combinando la técnica de rough sets con las redes neuronales y fuzzy logic.

La teoría de rough sets se compara con las técnicas de Aprendizaje de máquina de la Inteligencia Artificial (tales como la Minería de datos o Data Mining y Generación de conocimiento-KDD) y los análisis estadísticos, particularmente los análisis discriminantes. Dentro de las aplicaciones en el campo empresarial para esta técnica se puede mencionar la predicción en el mercado accionario para el uso de los inversionistas en sus decisiones de compra y venta, predicción del fracaso financiero, database marketing (búsqueda de patrones en el consumo de los clientes).

1.5 Descripción de las herramientas y técnicas a utilizar.

1.5.1 Método Delphi de la consulta de expertos.

En este trabajo se utilizará la técnica de Consulta a Expertos para asegurar que el conocimiento que se obtenga esté respaldado por la opinión de los especialistas, lográndose con ello una mayor solidez y rigor científico en la investigación.

El método de expertos se basa en la consulta a personas a las que se supone que tienen conocimientos elevado de la materia que se va a tratar en la organización donde desarrolla su labor. Estas personas exponen sus ideas y finalmente se redacta un informe en el que se indican cuáles son, en su opinión, las posibles alternativas que se tendrán en el futuro.

El método de expertos ideal sería aquel que extrajese los beneficios de la interacción directa y eliminase sus inconvenientes. Esta intenta ser la filosofía de la metodología Delphi; que se empleará como apoyo para el análisis estadístico el software Statgraphics, tratándose de reunir la mayor cantidad posible de elementos que puedan aportar datos significativos.

Este método presenta tres características fundamentales [18]:

- ✓ Anonimato: Ningún experto conoce la identidad de los otros que componen el grupo de debate.
- ✓ Iteración y realimentación controlada: La iteración se consigue al presentar varias veces el mismo cuestionario. Como además se van presentando los resultados obtenidos con los cuestionarios anteriores, se consigue que los expertos vayan conociendo los distintos puntos de vista y puedan ir modificando su opinión si los argumentos presentados les parecen más apropiados que los suyos.
- ✓ Respuesta del grupo en forma estadística: La información que se presenta a los expertos no es sólo el punto de vista de la mayoría, sino que se presentan todas las opiniones indicando el grado de acuerdo que se ha obtenido.

1.5.2 Statgraphics.

Statgraphics se utilizará para dar tratamiento a la Consulta a Expertos. Se seleccionó por ser una herramienta para análisis estadístico dirigida por menús que integra una gran variedad de gráficos de alta resolución., es sencillo de aprender y utilizar gracias a su diseño intuitivo que facilita la realización de los diversos análisis.

La estructura de Statgraphics aporta al usuario información sobre relaciones entre datos así como gráficos a color de gran calidad, aún 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. [19]

1.5.3 WEKA.

WEKA es el acrónimo de Waikato Environment for Knowledge Analysis y nace del esfuerzo de un grupo de investigadores del Machine Learning Laboratory de la Universidad de Waikato en Nueva Zelanda, como software de código abierto bajo los términos de la GNU GPL y es un banco completo de herramientas para aprendizaje automático y minería de datos.

Weka es una colección de algoritmos de aprendizaje automático para la extracción de datos. Los algoritmos pueden ser aplicados directamente a un conjunto de datos o llamarlos desde su propio código Java. Weka contiene herramientas para los datos de pre-procesamiento, clasificación, regresión, clustering, reglas de asociación, y la visualización. Está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla.

La licencia de Weka es GPL*, lo que significa que este programa es de libre distribución y difusión. Además, ya que Weka está programado en Java, es independiente de la arquitectura, ya que funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible. [16]

Con este software se realizará el procesamiento de los datos contenidos en la Base de Hechos, se obtendrán un conjunto de reglas a partir de las que estará soportado el sistema. Se escogió este software porque ya trae algoritmos implementados, con los que se puede generar el AD. Además es multiplataforma (Java) y tiene cuatro interfaces de usuario distintas, permitiendo múltiples funcionalidades.

1.5.4 ROSE2.

Rose2 (Rough Sets Data Explorer) es un software de la aplicación de elementos básicos de la teoría de conjuntos aproximados y técnicas de descubrimiento de reglas. Se creó en el Laboratorio de Intelligent Decision Support Systems, del Instituto de Ciencias de Computación en Poznan.

Se escogió este software porque el sistema consta de un entorno gráfico fácil de usar, cómodas herramientas para la exploración y análisis de datos y una serie de módulos de cálculo. Los módulos son independientes de la plataforma y se puede compilar para diferentes sistemas operativos, incluyendo Linux. El entorno actúa como una plantilla en todos los módulos de cálculo. Es fácil de añadir nuevos módulos para el sistema siendo esta una característica importante. Esto garantiza una mayor capacidad de expansión del sistema en el futuro.

Entre las características que ofrecen los módulos de cálculo se encuentran:

- ✓ Validación de datos y procesamiento preliminar.
- ✓ Discretización automática de atributos de valor constante.
- ✓ Estimación cualitativa de la capacidad de la condición de los atributos de la clasificación aproximada de los objetos.
- ✓ Conjunto de datos basado en el análisis sobre la relación de semejanza.
- ✓ La aplicación de un conjunto de reglas de decisión para clasificar los objetos con un nuevo clasificador de distancia.

1.5.5 Lenguaje de programación.

El C# se utilizará para la realización de la aplicación. Existen muchos otros lenguajes de programación, se seleccionó este por presentar una serie de ventajas que lo hacen muy atractivo, tanto para su uso profesional como para el aprendizaje de la programación. Es un lenguaje de paradigma orientado a objetos. Cumple con las expectativas que requiere la aplicación a implementar y ya se conoce este lenguaje por lo que no es necesario aprender uno nuevo para poder desarrollar este trabajo, siendo esta una ventaja en cuanto al empleo de tiempo que se dispone.

CAPÍTULO II APLICACIÓN DEL CONOCIMIENTO

2.1 Introducción

En este capítulo se hace un análisis de la consulta de expertos con el propósito de validar la importancia y necesidad del inicio de la investigación, utilizando técnicas del método Delphi. Se presenta una plantilla que recoge los parámetros correspondientes al análisis de costo, además de acuerdo con la información obtenida, se confecciona la base de datos vinculada al proyecto SCADA.

Se dan los pasos necesarios para la aplicación de la teoría de los conjuntos aproximados, se analiza el conjunto de datos para obtener el sistema de información compuestos por atributos y objetos donde a cada atributo de le asigna un valor del dominio.

2.2 Método Delphi de la consulta de expertos.

Se consultaron especialistas con 4 o más años de experiencia en análisis de costo presentándole un cuestionario de preguntas vinculando dos variables de análisis de costo. Un ejemplo de estas variables es: Gestión de Costo (GC) y Costo de Producción (CP), donde GC-CP se refería a la vinculación de la primera variable con la segunda, las respuestas de los expertos están dadas por los siguientes criterios: Bastante Relación, Relación, Poca Relación, Muy Poca Relación y Ninguna Relación.

2.2.1 Cuestionario aplicado a los Expertos.

El cuestionario fue aplicado a 12 expertos graduados en ingeniería industrial, eléctrica e informática, licenciatura en economía o contabilidad, licenciatura en matemática. El objetivo fue llegar a un acuerdo final de cuáles son las relaciones entre las variables y sus dependencias con el propósito de validar la importancia y necesidad del tema de investigación.

2.2.1.1 Cuestionario de preguntas.

Marque con una X la respuesta que usted considere correcta a la siguiente interrogante. ¿Las variables que se muestran a continuación guardan relación?

Bastante Relación (BR). Relación (R). Poca Relación (PR).

Muy Poca Relación (MR). Ninguna Relación (NR).

Relación entre Variables	BR	R	PR	MP	N
Gestión de costos - Costo de producción.					
Gestión de costos - Gestión de costo de un proyecto de SW.					
Gestión de costos - Rentabilidad del producto.					
Gestión de costos - Factibilidad de un proyecto de SW.					
Gestión de costos - Toma de decisiones.					
Gestión de costos - Control de costo.					
Gestión de costos - Estimación de costos.					
Gestión de costos - Presupuesto de costo.					
Costo de producción - Rentabilidad del producto.					
Costo de producción - Función de costo.					
Costo de producción - Toma de decisiones.					
Costo de producción – Costo del Estudiante.					
Costo de producción - Salario de profesores.					
Costo de producción - Costo de mano de obra directa.					
Costo de producción - Servicios telefónicos.					
Costo de producción - Servicios de conectividad.					
Costo de producción - Consumo eléctrico.					
Costos de Producción - Costo de insumo.					
Toma de decisiones - Rentabilidad del producto.					
Toma de decisiones - Estimación de costos.					
Toma de decisiones - Árbol de decisión.					
Inteligencia Artificial - Conjuntos Aproximados.					
Conjuntos Aproximados - Función de Costo.					
Costo de producción - Gestión de costos.					
Rentabilidad del producto - Gestión de costos.					
Toma de decisiones - Gestión de costos.					
Rentabilidad del producto - Costo de producción.					
Toma de decisiones - Costo de producción.					
Consumo eléctrico - Costo de producción.					
Salario de profesores - Costo de producción.					
Toma de decisiones - Conjuntos Aproximados.					
Conjuntos Aproximados - Toma de decisiones.					

Tabla 2: Cuestionario de preguntas.

2.2.1.2 Resultados obtenidos.

De las respuestas dadas al cuestionario anterior se obtuvieron las dependencias existentes entre las variables. Debido a que las opiniones de los expertos pueden ser distintas se da como resultado final aquel que represente un mayor por ciento de coincidencia en los criterios que se emitieron. A continuación se mostrarán los resultados obtenidos.

Variable 1	Variable 2	Relación
Gestión de costos	Costo de producción	Bastante Relación
Gestión de costos	Gestión de costos de proyectos de SW	Relación
Gestión de costos	Rentabilidad del producto	Bastante Relación
Gestión de costos	Factibilidad de un proyecto de SW	Relación
Gestión de costos	Toma de decisiones	Bastante Relación
Gestión de costos	Control de costo	Relación
Gestión de costos	Estimación de costos	Relación
Gestión de costos	Presupuesto de costo	Relación
Costo de producción	Rentabilidad del producto	Bastante Relación
Costo de producción	Función de costo	Relación
Costo de producción	Toma de decisiones	Bastante Relación
Costo de producción	Costo de estudiantes	Relación
Costo de producción	Salario de profesores	Bastante Relación
Costo de producción	Costo de mano de obra directa	Relación
Costo de producción	Servicios telefónicos	Relación
Costo de producción	Servicios de conectividad	Relación
Costo de producción	Consumo eléctrico	Bastante Relación
Costo de producción	Costo de insumo	Relación
Toma de decisiones	Rentabilidad del producto	Relación
Toma de decisiones	Estimación de costos	Relación
Toma de decisiones	Árbol de decisión	Relación
Inteligencia Artificial	Conjuntos Aproximados	Relación
Conjuntos Aproximados	Función de Costo	Relación
Costo de producción	Gestión de costos	Bastante Relación
Rentabilidad del producto	Gestión de costos	Bastante Relación
Toma de decisiones	Gestión de costos	Bastante Relación
Rentabilidad del producto	Costo de producción	Bastante Relación
Toma de decisiones	Costo de producción	Bastante Relación
Consumo eléctrico	Costo de producción	Bastante Relación
Salario de profesores	Costo de producción	Bastante Relación

Toma de decisiones	Conjuntos Aproximados	Bastante Relación
Conjuntos Aproximados	Toma de decisiones	Bastante Relación

Tabla 3: Resultados obtenidos.

Una vez obtenidos los resultados de los expertos se le asigna un identificador a cada variable.

Variables	
Gestión de costos (GC).	Servicios de conectividad (SC).
Gestión de costo de un proyecto de SW (GP).	Consumo eléctrico (CE).
Toma de decisiones (TD).	Costo de insumo (CI).
Estimación de costos (EC).	Inteligencia Artificial (IA).
Presupuesto de costo (PC).	Conjuntos Aproximados (CA).
Costo de producción (CP)	Función de Costo (FC).
Costo de mano de obra directa (CM).	Costo del Estudiante (CD).
Servicios telefónicos (ST).	Salario de profesores (SP).
Árbol de decisión (AD).	Control de costo (CC).
Factibilidad de un proyecto de SW (FP).	Rentabilidad del producto (RP).

Tabla 4: Identificador de cada variable.

En la figura 4 que se muestra a continuación GC, GP, TD, EC, PC, CP, CM, ST, AD, FP, SC, CE, CI, IA, CA, FC, CD, SP, CC y RP son puntos que reciben el nombre de vértices y las líneas en un sentido (o dos), arcos. Las saetas de doble flecha representan la existencia de mucha relación entre los dos vértices que une, los cuales representan las variables, mientras que la saeta simple representa que existe relación.

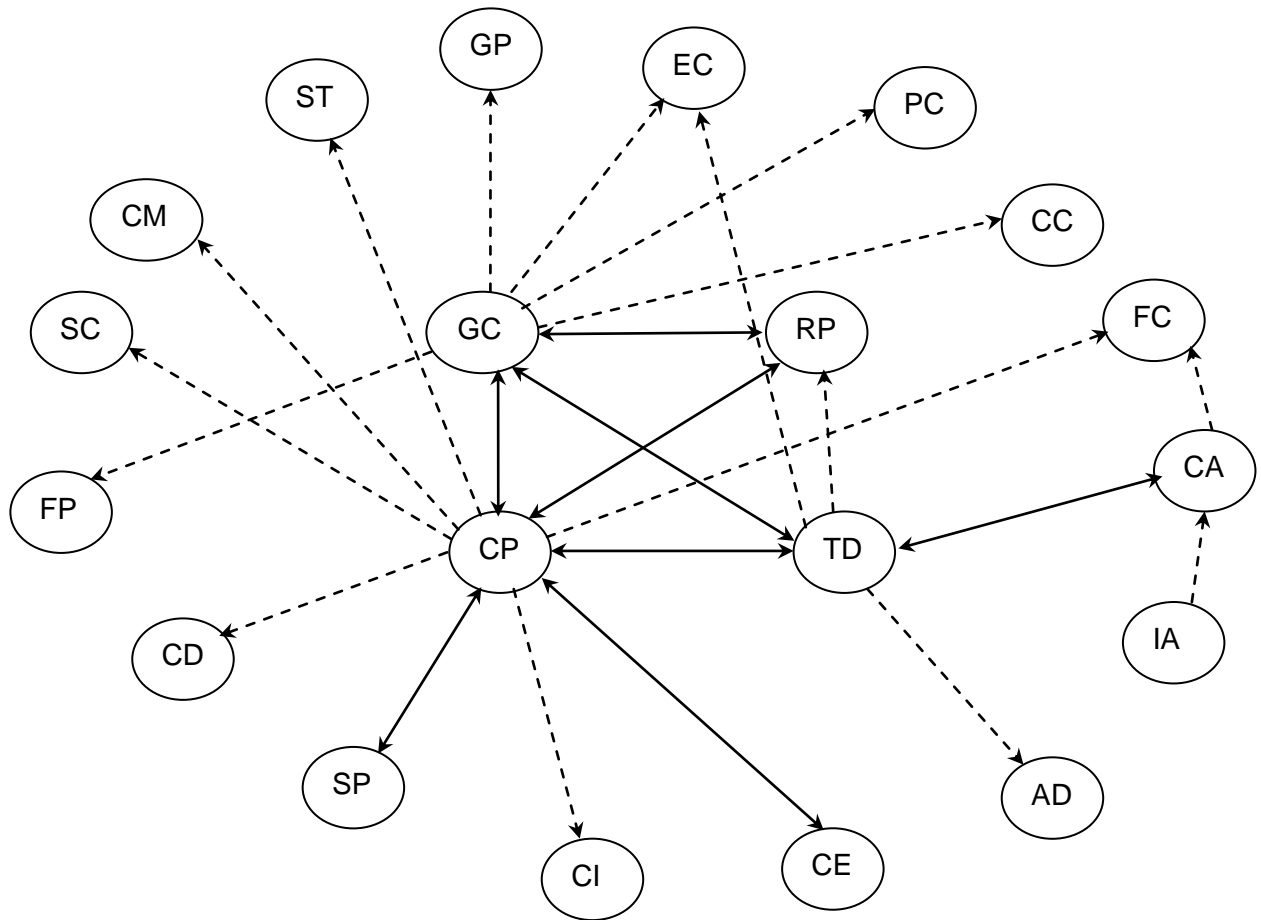


Figura 4: Grafo de relaciones entre las variables.

Fuente: Elaboración propia.

En la tabla 5 se presentan los resultados obtenidos en el cuestionario aplicado a los expertos, según la relación entre las variables, donde se le asigna a: ninguna relación: 0, muy poca relación: 0, poca relación: 0, relación: 1, bastante relación: 1.

Matriz Booleana																				
	GC	GP	TD	EC	PC	CP	CM	ST	AD	FP	SC	CE	CI	IA	CA	FC	CD	SP	CC	RP
GC	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1
GP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TD	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1
EC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CP	1	0	1	0	0	0	1	1	0	0	1	1	1	0	0	1	1	1	0	1
CM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CE	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
CA	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
FC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RP	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 5: Matriz Booleana.

En la tabla 6 se muestra el análisis de los poderes (tipo A o tipo B) de cada vértice que se obtienen a partir de los resultados que fueron mostrados en la Matriz Booleana, en la cual la suma de las filas nos dará el número total de variables que directamente necesitan de otras (ejemplo CP necesita de GC) y representan el poder de tipo A, mientras que la suma de las columnas nos dará el número de poderes del tipo B; las variables que tengan el mismo valor tendrán el mismo poder (peso). Al hacer este análisis se logra el orden de las variables con respecto al valor de su poder donde el vértice con mayor poder es la variable con más dependencia y de la que más dependen las otras.

Tabla de poderes.			
	PA	PB	PA + PB
GC	8	3	11
GP	0	1	1
TD	6	3	9
EC	0	2	2
PC	0	1	1
CP	11	5	16
CM	0	1	1
ST	0	1	1
AD	0	1	1
FP	0	1	1
SC	0	1	1
CE	1	1	2
CI	0	1	1
IA	1	0	1
CA	2	2	4
FC	0	2	2
CD	0	1	1
SP	1	1	2
CC	0	1	1
RP	2	3	5

Tabla 6: Tabla de poderes.

Según el criterio de los expertos las variables con mayor peso son CP, GC y TD correspondientes a: Costo de producción, Gestión de costos y Toma de decisiones.

2.3 Análisis estadístico de la información.

Después que se procesa el cuestionario aplicado a los expertos, se procede al análisis correspondiente con ayuda del asistente estadístico Statgraphics, para validar que los resultados obtenidos en el cuestionario sean lo más confiable posible.

Se encontró la media aritmética de los criterios de los mismos por cada elemento, determinando el coeficiente de variación por medio del cálculo de la desviación típica S.

Si $S > 1$ Se rechaza el criterio de los expertos y se realiza una nueva ronda de preguntas.

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

Si $S = 1$ Significa un acuerdo total entre los expertos.

El análisis estadístico para cada una de las combinaciones entre variables quedó de la siguiente manera:

Variables	Cantidad	Promedio	Varianza	Desviación Típica	Mínimo	Máximo	Rango
GC - CP	11	472,727	0,21818	0,4671	4,0	5,0	1,0
GC - GP	11	490,909	0,09091	0,30515	4,0	5,0	1,0
GC - RP	12	416,667	124,244	111,464	1,0	5,0	4,0
GC - FP	12	391,667	208,333	144,338	1,0	5,0	4,0
GC - TD	12	4,75	0,20455	0,45227	4,0	5,0	1,0
GC - CC	12	4,5	0,81818	0,90453	2,0	5,0	3,0
GC - EC	11	481,818	0,16364	0,40452	4,0	5,0	1,0
GC - PC	11	481,818	0,16364	0,40452	4,0	5,0	1,0
CP - RP	11	481,818	0,16364	0,40452	4,0	5,0	1,0
CP - FC	12	4,5	0,27273	0,52222	4,0	5,0	1,0
CP - TD	12	458,333	0,26515	0,51493	4,0	5,0	1,0
CP - CD	12	4,5	0,81818	0,90453	2,0	5,0	3,0
CP - SP	12	433,333	133,333	11,547	1,0	5,0	4,0
CP - CM	11	472,727	0,21818	0,4671	4,0	5,0	1,0
CP - ST	12	3,5	190,909	13,817	1,0	5,0	4,0
CP - SC	12	4,0	109,091	104,447	1,0	5,0	4,0
CP - CE	12	4,25	0,204545	0,452267	4,0	5,0	1,0
CP - CI	12	433333	0,42424	0,65134	3,0	5,0	2,0
TD - RP	12	4,5	0,81818	0,90453	2,0	5,0	3,0
TD - EC	12	441,667	0,44697	0,66856	3,0	5,0	2,0
TD - AD	12	433,333	133,333	11,547	1,0	5,0	4,0
IA - CA	9	422,222	0,69444	0,83333	3,0	5,0	2,0
CA - FC	10	3,9	0,76667	0,8756	2,0	5,0	3,0
CP - GC	11	472,727	0,21818	0,4611	4,0	5,0	1,0
RP - GC	12	416,667	124,244	111,464	1,0	5,0	4,0
TD - GC	12	4,75	0,20455	0,45227	4,0	5,0	1,0
RP - CP	11	481,818	0,16364	0,40452	4,0	5,0	1,0
TD - CP	12	458,333	0,26515	0,51493	4,0	5,0	1,0
CE - CP	12	4,29	0,20455	0,45227	4,0	5,0	1,0
SP - CP	12	433,333	133,333	11,547	1,0	5,0	4,0
TD - CA	12	458,333	0,26515	0,51493	4,0	5,0	1,0
CA - TD	12	458,333	0,26515	0,51493	4,0	5,0	1,0

Tabla 7: Análisis estadístico entre las variables.

Se aceptó el criterio de los expertos porque al calcular la desviación típica para cada una de las relaciones existentes entre las variables, la misma de manera general está por debajo de uno, 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.

2.3.1 Gráficos Obtenidos.

Las siguientes gráficas representan la relación entre dos variables. Se muestra la relación de la concordancia de criterios de los expertos en cuanto a las respuestas en una misma pregunta, el eje vertical muestra la cantidad de expertos entrevistados y el eje horizontal muestra los criterios que se tuvieron en cuenta para darle valor a las variables.

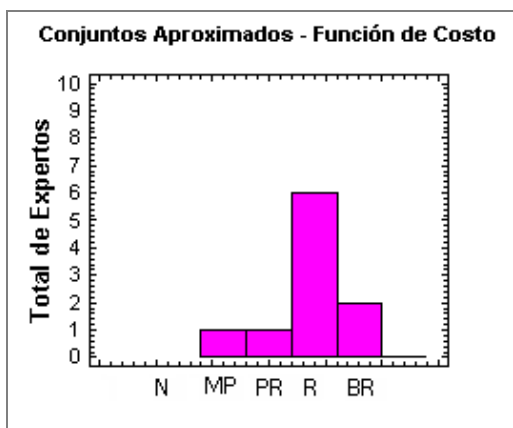


Figura 5: Gráfico de la relación de las variables CA-FC

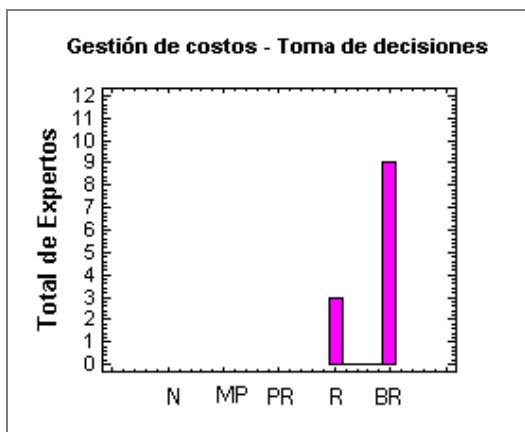


Figura 6: Gráfico de la relación de las variables GC-TD.

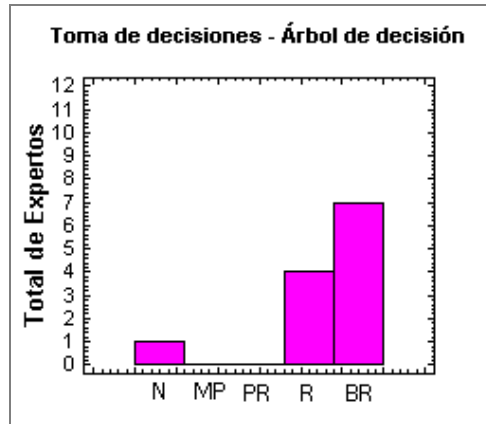


Figura 7: Gráfico de la relación de las variables TD-AD.

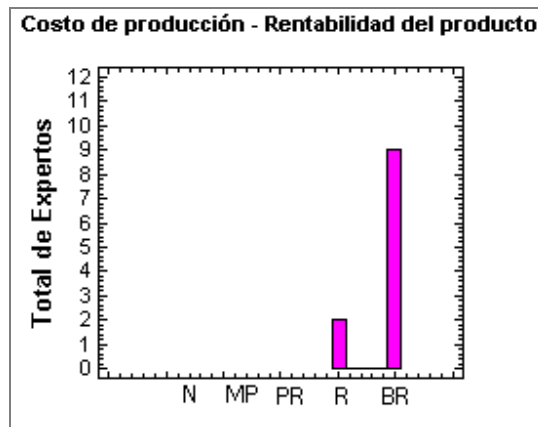


Figura 8: Gráfico de la relación de las variables CP-RP.

En el gráfico siguiente se hace una representación de la cantidad de frecuencia con que los expertos, al responder la pregunta del cuestionario vieron que existía importancia entre las relaciones de las variables.

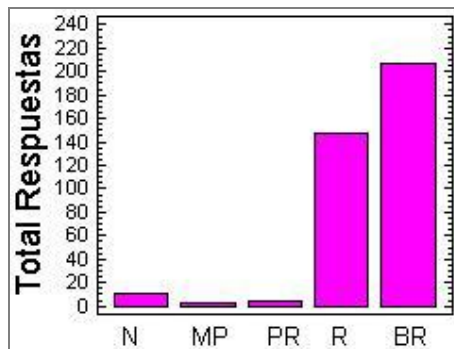


Figura 9: Gráfico de la cantidad de frecuencia con que los expertos respondieron.

Esta misma representación se muestra a continuación mediante la siguiente gráfica de los por ciento de la concordancia de los expertos según cada criterio.

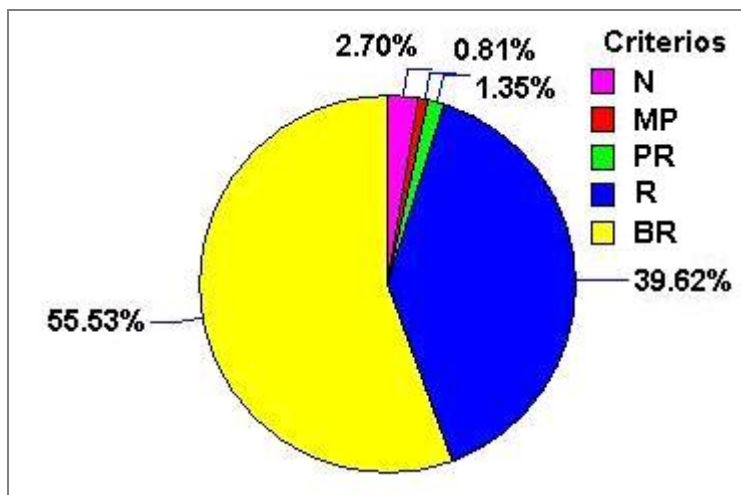


Figura 10: Representación del por ciento de la concordancia de los expertos.

2.4 Elaboración de la Base de Hechos.

En la realización de este trabajo se utiliza una Bases de Hechos perteneciente a los datos de los gastos en los que incurre el proyecto productivo SCADA. Para su confección se consultaron varias plantillas que se usan para el análisis de costo. Las mismas fueron obtenidas a través de la investigación realizada al Departamento de Ciencias Empresariales (DCE) y a la Alternativa Bolivariana para la Exportación de las Américas (ALBET).

2.4.1 Propuesta del Departamento de Ciencias Empresariales.

En la figura siguiente se muestra una síntesis de la plantilla del DCE con los principales gastos por elementos y subelementos, agrupándolos de acuerdo a si se incurren o no en el proyecto, quedando de la siguiente manera:

	A	B	C	D	E
1		Enero		Febrero	
2	Elementos y subelementos de gasto.	CUC	MN	CUC	MN
3	Total	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
4	Materias primas y materiales	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
5	Insumos informáticos				
6	Materiales auxiliares				
7	Materiales de oficina				
8	Otros gastos materiales				
9	Portadores energéticos	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
10	Gasolina motor regular.				
11	Gasolina motor especial				
12	Diesel para plantas eléctricas				
13	Diesel para transporte				
14	Otros combustibles				
15	Electricidad				

Figura 11: Plantilla para el control de los elementos y subelementos de gastos.

2.4.2 Plantilla proporcionada por ALBET.

El departamento de ALBET facilitó un Excel dividido por varias pestañas que representan los componentes con los se realiza el análisis de costo, en la figura que a continuación se muestra se presenta una de ellas.

	A	B	C	D
1	Presupuesto para ejecutar la solución			
2	0.00	0.00	Mes 1	Mes 2
3	PRESUPUESTO EFECTIVO TOTAL	0.00	0.00	0.00
4	MATERIAS PRIMAS Y MATERIALES	0.00	0.00	0.00
5	Materias Primas	0.00		
6	Materiales Auxiliares	0.00		
7	Partes y Piezas de Repuesto	0.00		
8	Útiles y Herramientas	0.00		
9	Materiales de Oficina	0.00		
10	Productos de Aseo y Limpieza	0.00		

Figura 12: Plantilla proporcionada por ALBET.

2.4.3 Modelo a utilizar.

Para obtener la Base de Hechos, se creó una nueva plantilla que vincula los datos más significativos de las dos propuestas anteriores.

	A	B	C	D	E
1		PRESUPUESTO EFECTIVO TOTAL	0.00	Proyecto 1	Proyecto 2
2	M	Materias Primas y Materiales	0.00	0.00	0.00
3	M1	Materias Primas	0.00		
4	M2	Materiales Auxiliares	0.00		
5	M3	Materiales de Oficina	0.00		
6	M4	Insumos Informáticos	0.00		
7	M5	Propiedad Intelectual	0.00		
8	M6	Licencia de uso de software	0.00		
9	M7	Otros Gastos Materiales	0.00		
10	P	Portadores Energéticos	0.00	0.00	0.00
11	P1	Gasolina de Motor Regular	0.00		
12	P2	Gasolina de Motor Especial	0.00		
13	P3	Diesel para Transporte	0.00		
14	P4	Otros Combustibles	0.00		
15	P5	Electricidad	0.00		
16	P6	Diesel para plantas eléctricas	0.00		

Figura 13: Modelo a utilizar.

2.4.3.1 Descripción de algunos elementos que contiene el modelo.

En el presente epígrafe se describen las variables utilizadas en la gestión de los costos en los proyectos productivos de la UCI como son: materias primas y materiales, impuestos y contribución, depreciación y amortización, salarios, portadores energéticos y otros gastos; todos estos elementos están divididos por subelementos, agrupándolos de acuerdo a si se incurren o no en el proyecto, aquí se dará una breve explicación de cada uno de ellos.

Variables	Presupuesto efectivo total	
M	Materias primas y materiales	
M1	Materias primas	Representa el valor de adquisición de las materias primas que se integran y consumen en el proceso de la producción o los servicios.

M2	Materiales auxiliares	Comprende los otros materiales que se utilizan en el proceso productivo o de servicio como complemento del mismo sin ser un producto básico.
M3	Materiales de Oficina	Representa el gasto de los materiales de oficina que se consumen en función de la actividad administrativa, incluye tóner y papelería, además agendas cuando son para uso del personal que trabajan en la entidad
M4	Insumos Informáticos	Comprende los importes pagados por los insumos necesarios para la actividad informática.
M5	Propiedad Intelectual	Representa los gastos por el registro legal del software que intervienen en cada uno de los proyectos.
M6	Licencia de uso de software	Representa el valor de adquisición de las licencias de uso, necesarias en la ejecución de un proyecto.
M7	Otros gastos materiales	Representa el valor consumido de otros materiales no contemplados anteriormente y se utilizan en el proceso productivo o de servicio así como en otras actividades administrativas. Incluye entre otros: maletines, baterías para medios electrónicos y de comunicaciones (ejemplo: batería de teléfonos inalámbricos, celulares, control remoto, beeper, etc.)
P	Portadores Energéticos	
P1	Gasolina de Motor Regular	Representa el valor de la gasolina regular consumida para ser utilizada en los medios de transporte u otro equipamiento
P2	Gasolina de Motor Especial	Representa el valor de la gasolina especial consumida para ser utilizada en los medios de transporte u otro equipamiento.
P3	Diesel para Transporte	Representa el valor del gas oíl consumido para producir energía, sea en el proceso de producción o

		servicios en la actividad principal; así como en la explotación de equipos de combustión interna en actividades no fundamentales, así como en los equipos de transporte.
P4	Otros Combustibles	Representa el valor de otros combustibles no considerados anteriormente y consumidos para producir energía, en el proceso de producción o prestación de servicios en la actividad fundamental, así como por otras actividades no fundamentales.
P5	Electricidad	Representa el consumo del fluido eléctrico suministrado y que se utiliza tanto en la actividad de producción o servicios prestados en la actividad fundamental, así como de apoyo a la actividad administrativa o comercial o de distribución.
P6	Diesel para plantas eléctricas	Gasto producido por el combustible utilizado para la obtención de energía eléctrica mediante plantas.
W	Salarios	
W1	Salario profesores.	Gasto de salario por el pago a los trabajadores.
W2	Descanso retribuido	Gasto por la remuneración a los trabajadores por vacaciones acumuladas.
W3	Estimulación salarial.	Gasto por la remuneración realizada a los trabajadores producto del desempeño en su puesto de trabajo.
W4	Vinculación	Gasto por la remuneración realizada producto de la vinculación a la producción del personal.
W5	Otros pagos de fuerza de trabajo	Otros pagos de la fuerza de trabajo no incluidos en los conceptos anteriores.
X	Impuestos y contribución.	
X1	Contribución a la Seguridad Social al	Gastos originados por el pago de la seguridad social a los trabajadores.

	14%	
X2	Subsidio pagado a corto plazo	Se registran los pagos que por enfermedad, maternidad u otro tipo d enfermedad se le retribuya a los trabajadores.
X3	Impuesto por la utilización de la fuerza de trabajo	Gasto por concepto de pago al trabajador en sustitución del salario, cuando este se enferma o accidenta.
D	Depreciación y Amortización.	
D1	Depreciación de Equipos de Informática	Representa el gasto por la transferencia paulatina en el proceso de producción, de la prestación de servicios, la comercialización, entre otras, de los medios y equipos informáticos. Se cargan a gastos mensualmente según las tasas establecidas.
D2	Depreciación de Muebles, Enseres y Similares	Representa el gasto por la transferencia paulatina en el proceso de producción, de la prestación de servicios, la comercialización, entre otras, de los muebles, enseres y similares. Se cargan a gastos mensualmente según las tasas establecidas.
D3	Depreciación de edificios y otras construcciones	Representa el gasto por la transferencia paulatina en el proceso de producción, de la prestación de servicios, la comercialización, entre otras, de los edificios y otras construcciones. Se cargan a gastos mensualmente según las tasas establecidas.
D4	Depreciación de equipos de comunicaciones	Gasto por concepto de Depreciación de Equipos de Comunicaciones.
D5	Depreciación de equipos de transporte	Representa el gasto por la transferencia paulatina en el proceso de producción, de la prestación de servicios, la comercialización, entre otras, de los equipos de transporte. Se cargan a gastos

		mensualmente según las tasas establecidas.
D6	Depreciación de equipos energéticos	Gasto por concepto de Depreciación de Equipos de Energéticos.
D7	Depreciación de otros activos fijos tangibles	Representa el gasto por la transferencia paulatina en el proceso de producción, de la prestación de servicios, la comercialización, entre otras, de los otros activos fijos tangibles. Se cargan a gastos mensualmente según las tasas establecidas.
D8	Amortización de activos fijos intangibles	Comprende la parte proporcional de aquellos gastos que se autoricen a diferir a corto o largo plazo por los conceptos de gastos de amortización de activos fijos intangibles, tales como patentes de productos, marcas, licencias, derecho de autor, software, etc.
D9	Amortización de activos diferidos a largo plazo.	Gasto por concepto de Amortización de Activos Diferidos a Largo Plazo.
Otros Gastos y Servicios		
H	Hospedaje	
H1	Hospedajes en el Extranjero	Gastos generados en el extranjero por el alojamiento del personal del proyecto en función del mismo.
H2	Hospedajes en el territorio nacional	Gastos generados por el alojamiento del personal del proyecto en función del mismo.
J	Pasaje	
J1	Pasajes Internacionales	Gastos generados por pasajes al extranjero del personal del proyecto en función del mismo.
J2	Pasaje en el territorio nacional	Gastos generados por pasajes relacionados con la transportación del personal del proyecto.
T	Viáticos	
T1	Viáticos en el territorio nacional	Gastos generados por el personal que labora en el territorio nacional en funciones del proyecto.

T2	Viáticos en el extranjero	Gastos generados por el personal que labora en el extranjero en funciones del proyecto.
G	Gastos Varios	
G1	Otros gastos por viajes	Representan los otros gastos de viajes no considerados anteriormente que se registran en función del trabajo. Se incluyen todos los gastos relacionados con la tramitación de documentos por los viajes al extranjero y cualquier otro gasto de viaje al extranjero, incluye trámite de pasaporte, visa, legalización de documentos, etc.
G2	Investigación y Desarrollo	Representa el gasto en que se incurre por el concepto de investigación y desarrollo de nuevas técnicas, sean para los sistemas de software, otros equipos de computación, electrónicos u otra índole.
G3	Documentación técnica y Capacitación	Representan los gastos en que se incurren por la adquisición de documentación especializada, incluye soporte digital. En lo referido a la capacitación, se registran en esta partida, todos los gastos asociados a la preparación del personal de los proyectos.
G4	Programas Computacionales	Representan los gastos en que se incurre por el desarrollo, modificación, actualización, etc., de programas, o sistemas computacionales.
G5	Auditorías y Supervisiones	Representan los gastos en que se incurren por el concepto de la contratación de auditorías o supervisiones. Se incluyen las destinadas a la certificación de Balances.
G6	Consultorías Técnicas	Gasto por el pago que se realiza por el servicio de un especialista por concepto de consultoría técnica.
G7	Impresión y reproducción	Gasto generado por el concepto de servicio de impresión y reproducción de documentos.

G8	Registros, marcas y patentes	Gasto por el pago que se realiza cuando la entidad crea un producto y registra la marca o patenta esta.
G9	Alimentación	Representan los gastos que por extensión del horario laboral, sea necesario incurrir para la alimentación de los trabajadores. O reforzamiento autorizado, por la labor que se encuentren desarrollando.
G10	Peaje	Gasto por el pago de peaje.
G11	Arrendamiento	Gasto por el pago que se realiza cuando la entidad crea un producto y registra la marca o patenta esta.
G12	Otros Arrendamientos	Gastos generados por el alquiler de activos a otras entidades para las funciones del proyecto.
O	Otros Gastos Comerciales	
O1	Publicidad y Promoción	Gastos generados por la comunicación de la oferta del proyecto en los posibles mercados relacionado con los medios de comunicación disponibles (radio, TV, revistas, sitios web). Se incluye además el gasto relacionado por anuncios y carteles.
O2	Ferias y Eventos en Cuba	Gastos generados por participación en ferias y eventos en el territorio nacional.
O3	Ferias y Eventos en el Extranjero	Gastos generados por participación en ferias y eventos en el extranjero.
O4	Gastos de Representación y Cortesía	Gastos generados por actividades de protocolo, atención a personalidades que visitan, a clientes y proveedores y a delegaciones extranjeras.
O5	Garantía a Clientes	Representa el gasto en que se incurre por el concepto de prestación de servicios o sustitución de medios o equipos adquiridos por los clientes, que presentan dificultades técnicas durante el período de garantía.
O6	Otros Tributos	Gastos por concepto de otros impuestos que la entidad paga y que no están descritos en las

		clasificaciones anteriores.
C	Comunicaciones	
C1	Teléfono y Fax (celulares)	Gastos generados por servicios de teléfono y fax.
C2	Servicios de telex, telégrafo y correos	Gastos generados por servicios de Telex, Telégrafo y Correos.
C3	Servicios de Conectividad	Representan los gastos en que se incurren por el concepto de poner a punto la conexión de uno o varios equipos electrónicos a redes internas y externas para que trabajen y permitan la transmisión y recepción de información a uno o varios puntos a través de medios técnicos de comunicación o procesador de datos informáticos mediante equipos electrónicos.
C4	Otros Servicios de Comunicaciones	En esta partida se agrupan los otros gastos en que se incurren por el concepto de comunicaciones, los cuales no quedan recogidos en los anteriormente relacionados, entre las que se incluye, la instalación de nuevas extensiones telefónicas, etc, incluye además trunking y beepers.
R	Servicios de Mantenimiento y Reparación	
R1	Mantenimiento y Reparación de Equipos Informáticos	Gastos generados por mantenimientos planificados o reparación de las PC y demás equipos informáticos del proyecto.
R2	Servicios de Mantenimiento y Reparación de Muebles y Enseres	Representa el gasto en que se incurre por el concepto de los mantenimientos y reparaciones menores, que no constituyen inversiones, en muebles, enseres y similares.
R3	Otros Servicios de Mantenimientos	Gastos generados por otros gastos relacionados a mantenimiento que no se incluyen en el mantenimiento y reparación a equipos informáticos.

S	Seguros	
S1	Seguros de Transporte	Gasto por el pago de las primas de seguro que se pagan a la entidad aseguradora por la ocurrencia de accidentes.
S2	Seguros de Almacenes	Gasto por el pago de primas de seguros que se pagan a la entidad aseguradora por el aseguramiento de almacenes.
S3	Seguros de Mercancías	Gasto por el pago de primas de seguros que se pagan a la entidad aseguradora por el aseguramiento de mercancías.
S4	Otros Seguros	Gasto por el pago de primas de otros seguros.
V	Servicios Varios	
V1	Servicios de Parqueo	Corresponde al importe que se paga por la estancia de los vehículos en un local o área destinada para el estacionamiento, en empresas u otras entidades.
V2	Asistencia Técnica	Gasto por la contratación de personas o entidades para la asistencia técnica.
V3	Otros Servicios Informáticos	Gasto por concepto de otros servicios informáticos no relacionados en las categorías anteriores.
V4	Traducciones	Incluye los gastos en que se incurre por el concepto de traducciones de libros técnicos, folletos, revistas, publicaciones o traducciones habladas.
V5	Tramites de Documentos	Comprenden los gastos incurridos por concepto de tramitación de documentos legales en entidades del país. Incluye los servicios de tramitación de documentos relacionados con los equipos técnicos, recipientes de gas a presión, vehículos: licencia operativa del transporte de vehículos, licencia de circulación, trámites de circulación por vías prohibidas, inspección técnica de los vehículos, otros trámites

		como pérdida de chapa, etc, (no incluye los trámites de documentos para los viajes al exterior).
V6	Suscripción y Publicaciones	Representan los gastos en que se incurren por la suscripción a revistas y otras publicaciones técnicas, incluye soporte digital.
V7	Fumigación	Constituye el gasto incurrido por el pago del servicio de fumigación.
V8	Otros Servicios	Gastos que se generan por el pago de otros servicios no relacionados en las clasificaciones anteriores.
V9	Servicios jurídicos o legales	Gasto por el pago realizado por los servicios jurídicos o legales prestados.

Tabla 8: Descripción de los elementos que contiene el modelo.

2.5 Base de hechos.

La Base de Hechos (BH) consta de datos que a pesar de ser ficticios se aproximan a la realidad (lógica económica). Para su elaboración se consultaron especialistas de ALBET, DCE y del proyecto SCADA donde a partir de sus experiencias se informó como desarrollar el análisis de costo para la confección de la BH, que una vez realizada se obtuvo su aceptación.

El análisis realizado consta de sesenta instancias (proyectos en los que está dividido el SCADA) que serán utilizadas como ejemplos de entrenamiento para la obtención del conocimiento. Contiene 12 atributos que fueron seleccionados a partir del modelo descrito anteriormente, además del costo de la producción, el precio de venta y rentabilidad del producto. A continuación se da una descripción de la manera en que se lleva a cabo el cálculo de estos.

El costo de producción se calcula sumando los atributos seleccionados en el modelo.

El Precio de venta se efectúa atendiendo al costo de producción, por eso se realiza de la siguiente manera:

Precio de venta=Costo de producción * 100 / (100- Margen de utilidad)

El Margen de utilidad por tratarse de una empresa productora de software para los proyectos de inversión extranjera se les está aplicando un 70% aunque pueden aumentar o disminuir dependiendo de los factores que influyen en el precio, por ejemplo: el valor que tenga el software existente a nivel mundial en el mercado.

A continuación se procede a efectuar el análisis de la rentabilidad del software; éste mide el éxito o fracaso que tiene el mismo. Se entiende como rentable aquella inversión en la que el valor de los rendimientos que proporciona es superior al de los recursos que utiliza.

La Base de Hechos fue creada en una plantilla Excel, donde se guardaron los valores de todos los campos para cada una de las instancias. ([Ver Anexo 1](#))

2.6 Sistema de decisión.

Una vez obtenida la base de hechos se define el sistema de decisión que va a estar conformada por un conjunto de objetos (representan proyecto o líneas de producción), cuenta con 14 atributos y uno de decisión (Rentabilidad del Producto), además del valor de los atributos que está conformado de 5 a 10 clases en dependencia del rango de valores que comprende cada uno de ellos.

La siguiente tabla muestra los identificadores de cada uno de los atributos:

Identificador	Atributos	Rango de valores
A1	Materias Primas y Materiales	1000-10 000
A2	Portadores Energéticos	200-7200
A3	Otros Gastos y Servicios	500-480000
A4	Otros Gastos Comerciales	200-2000
A5	Otros Tributos	200-5000
A6	Comunicaciones	100-500
A7	Servicios de Mantenimiento y Reparación	60-500
A8	Servicios inmobiliarios y de Protección	60-500
A9	Seguros	25-200
A10	Servicios Varios	60-500
A11	Salarios	1500-72000
A12	Impuestos y contribución.	75-3600
A13	Costo de producción	3900-582000
A14	Precio de Venta	13260-1940000

Tabla 9: Identificador para cada atributo.

A continuación se muestra una tabla de los atributos, con los posibles valores que pueden tomar, así como el orden en el que aparecen en la Base de Hechos.

Atributo	Valor del dominio
A1	1000-2125, 2126-3250, 3251-4375, 4376-5500, 5501-6625, 6626-7750, 7751-8875, 8876-10000
A2	200-1075, 1076-1950, 1951-2825, 2826-3700, 3701-4575, 4576-5450, 5451-6325, 6326-7200
A3	500-48450, 48451-96400, 96401-144350, 144351-192300, 192301-240250, 240251-288200, 288201-336150, 336151-384100, 384101-432050, 432051-480000
A4	200-500, 501-800, 801-1100, 1101-1400, 1401-1700, 1701-2000
A5	200-800, 801-1400, 1401-2000, 2001-2600, 2601-3200, 2601-3200, 3201-3800, 3801-4400, 4400-5000
A6	100-180, 181-260, 261-340, 341-420, 421-500
A7	60-148, 149-236, 237-324, 325-412, 413-500
A8	60-149, 150-237, 238-325, 326-413, 414-500
A9	25-60, 61-95, 96-130, 131-165, 166-200
A10	60-147, 148-235, 236-323, 324-411, 412-500
A11	1500-9333, 9334-17166, 17167-24999, 25000-32832, 32833-40665, 40666-48498, 48499-56331, 56332-64164, 64165-72000
A12	75-578, 579-1081, 1082-1584, 1585-2087, 2088-2590, 2591-3094, 3095-3600
A13	3900-61710, 61711-119520, 119521-177330, 177331-235140, 235141-292950, 292951-350760, 350761-408570, 408571-466380, 466381-524190, 524190-582000
A14	13260 -205934, 205935-398608, 398609-591282, 591282-783956, 783957-967630, 967631-1169304, 1169305-1361978, 1361979-1554652, 1554653-1747362, 1747363-1940000
D	YES, NO

Tabla 10: Valores de la Base de Hechos.

La tabla 11 describe algunos de los elementos del sistema de decisión.

	A1	A2	A8	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 11: Ejemplo de la representación del sistema de decisión.

Al sistema de decisión anterior se le aplica la selección de rasgos con el propósito obtener un subconjunto de atributos describa el sistema como si se tratara del conjunto completo. Esto quiere decir que el proceso se centra en encontrar los atributos más importantes del sistema, observándose que todos son relevantes.

CAPÍTULO III. PROPUESTA DE SOLUCIÓN

3.1 Introducción

En este capítulo se realiza la obtención del conocimiento utilizando los software WEKA y ROSE. Se obtienen un conjunto de reglas que clasifican los objetos de la base de hechos, las cuales constituyen la base de la aplicación que permite conocer si un proyecto es rentable o no. Se mostrará la manera en que fue creado el sistema para el análisis de costo del proyecto SCADA, así como las interfaces que lo conforman, sus principales características y algunos resultados obtenidos de las pruebas aplicadas al mismo.

3.2 Procesamiento de datos.

El procesamiento de datos es el principal paso cuando se está tratando de obtener conocimiento de conjuntos de datos. En el caso de este trabajo el proceso consta de dos fases principalmente: la preparación de los datos a ser procesados y el procesamiento en sí. Una vez que la Base de Hechos fue creada y revisada se procedió al análisis de los datos utilizando los software ROSE y WEKA.

El ROSE es un software de técnicas de descubrimiento de reglas y el WEKA es una colección de algoritmos de aprendizaje automático para la extracción de datos, se utilizan ambos software para comparar el resultado de la calidad de la clasificación y en dependencia de dicho valor generar las reglas de producción.

3.2.1 Preparación de los Datos en el ROSE.

La primera operación que se realizó fue convertir las Bases de Hechos en un fichero extensión *.isf* (Information System File) que es utilizada en el ROSE para almacenar la información de los datos del sistema, por ejemplo, declaración de atributos y la definición de los objetos. Para esto se pasaron todos los datos de la Base de Hechos a un archivo Wordpad con nombres FicheroROSE y se guardó con la extensión especificada. En el fichero a cada atributo se le especifica los valores que pueden tomar entre corchetes. A continuación se muestra como quedó el fichero *.isf*

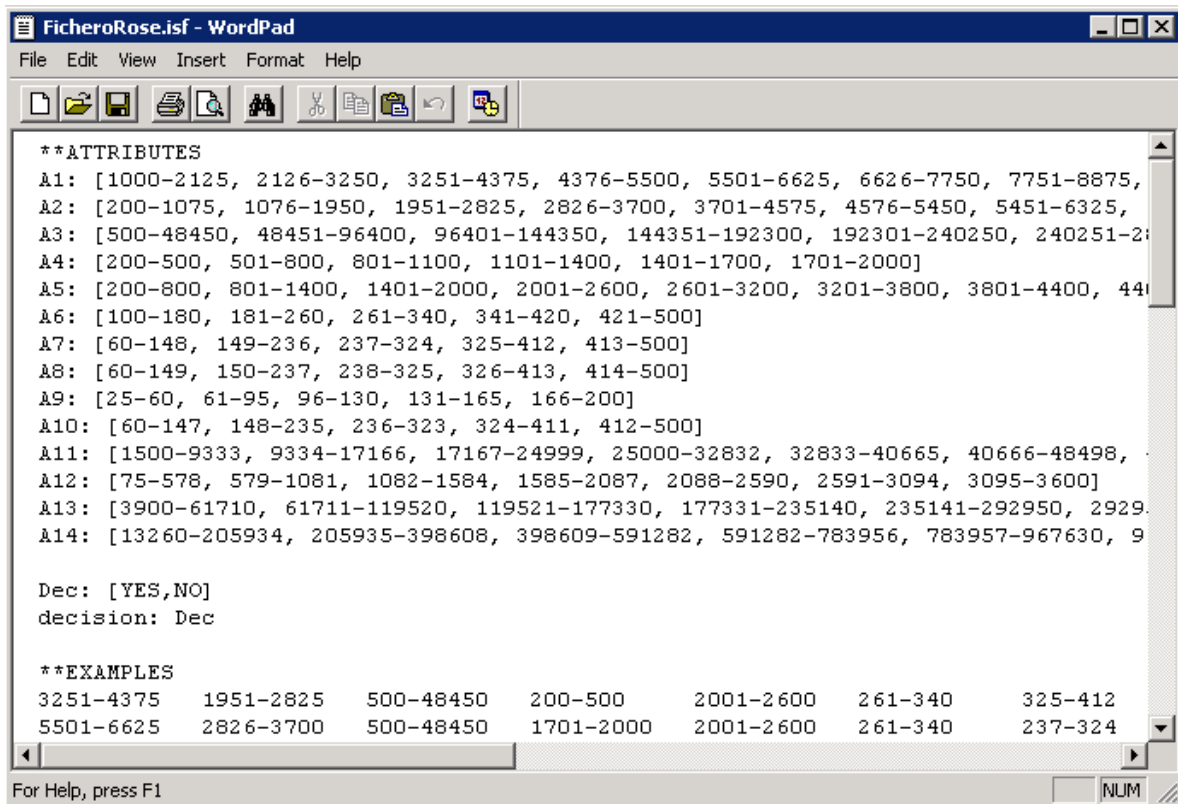


Figura 14: Archivos .isf

3.2.1.1 Procesamiento de la Base de Hechos y resultados obtenidos.

Después de creado el archivo *.isf* se procedió al procesamiento de los datos en el software ROSE2. El primer paso del procesamiento que se realizó fue cargar el fichero de la Base de Hechos en el software, donde por cada objeto se visualizan las características de los atributos. Seguidamente se procedió a la vista de las aproximaciones la cual quedó de la siguiente manera:

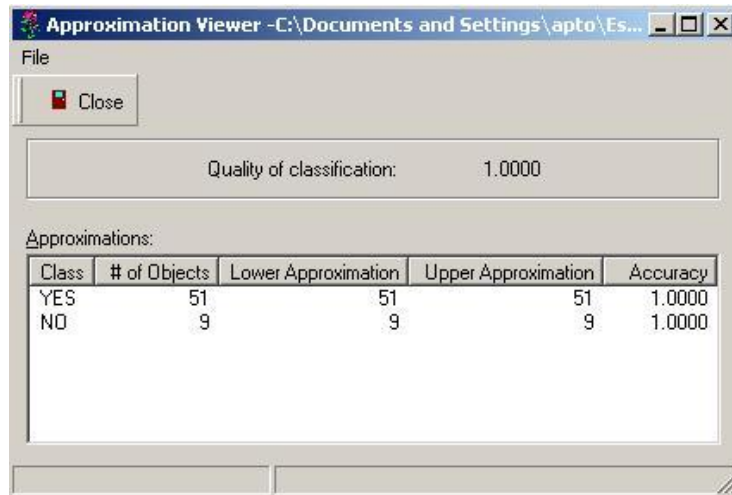


Figura 15: Aproximaciones.

En la figura 15 se puede observar que la calidad de clasificación es de 1; representa que no hay ningún objeto dudoso y el sistema es consistente.

Una vez obtenidas las aproximaciones se procedió a clasificar los resultados que se muestra a continuación:

```

=====
Classification Results
=====
Test: k-fold cross validation
Number of passes: 10
Majority threshold: 21.00%
Minimum similarity: 50.00%
Partially matched rules: all
Rule support: strength
=====
Confusion Matrix (sum over 10 passes)
-----
  YES NO None
YES 48 2 1
NO 0 2 7
=====
Average Accuracy [%]
-----
  Correct Incorrect None
Total 83.33 +- 14.91 3.33 +- 6.67 13.33 +- 12.47
YES 94.17 +- 11.81 4.17 +- 8.54 1.67 +- 5.00
NO 15.00 +- 32.02 0.00 +- 0.00 45.00 +- 47.17
=====

```


Como se mostró en la matriz de confusión sólo están mal clasificados 2 elementos de un total de 60 que pertenecen a la clase si lo que representa un 94.17% de la información bien clasificada.

3.2.2 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 que soporta el software a utilizar. Para esto se pasaron todos los datos de la Base de Hechos a un archivo Wordpad con nombres FicheroWEKA 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 a la hora de 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 entre llaves. A continuación se muestra como quedó el fichero *.arff*:

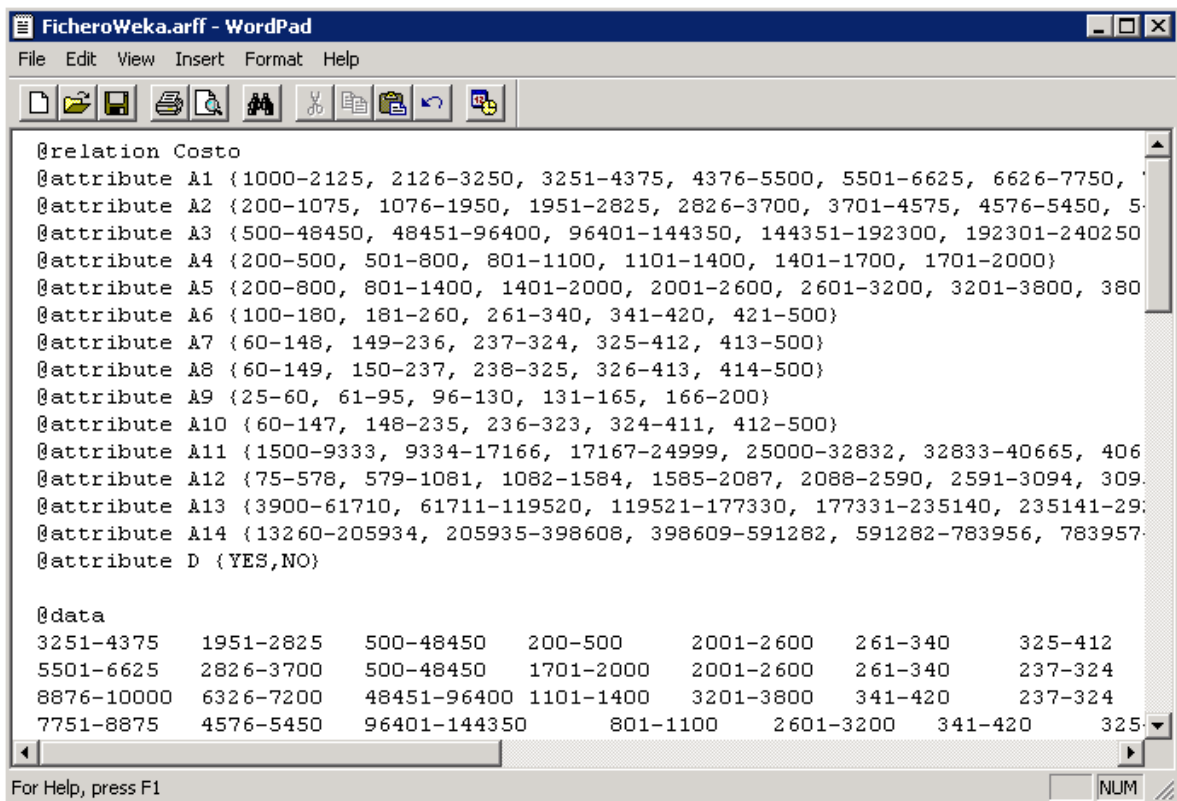


Figura 16: Archivos *.arff*

3.2.2.1 Procesamiento de la Base de Hechos y resultados obtenidos.

Una vez creado el archivo **.arrf** se comienza el procesamiento de los datos en 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. [\(Ver Anexo 4\)](#) A esta primera acción se le llama preprocesamiento. En este caso no fue necesario discretizar los atributos, que no es más que convertir los valores numéricos en nominales, ya que todos tienen valor nominal. [\(Ver Anexo 2\)](#)

3.3 Árbol de decisión.

El software WEKA permite con otros algoritmos como el ADTree, FT y J48 visualizar el AD, pero con el ID3 no ocurre así. Por esta razón a partir de los resultados obtenidos del procesamiento de datos de las Bases de Hechos se construyó el AD de la misma, para lograr con esto un mejor entendimiento del conocimiento que se alcanzó en el procesamiento.

Para seleccionar el AD se analizar los resultados utilizando 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)
ADTree	100	0.02
BFTree	98.3333	0.16
DecisionStump	80	0.02
FT	90	0.06
Id3	100	0.02
J48	85	0.02
J48graft	85	0.06
RandomTree	80	0.02
REPTree	85	0.02
SimpleCart	90	0.06

Tabla 12: Comparación entre algunos clasificadores de tipo árbol.

Como puede observarse en la tabla 12, los algoritmos de mayor calidad de clasificación fueron Id3 y el ADTree, por eso se decidió utilizarlos

Para realizar el procesamiento de los datos se seleccionó la pestaña *Classify* y dentro de ella se eligió el

clasificador *Trees* que es el correspondiente a los AD, y se eligió el algoritmo Id3 para el procesamiento ([Ver Anexo 3](#)). Una vez llevado a cabo todo este proceso se aplicó el procesamiento obteniéndose el siguiente resultado:

=== Classifier model (full training set) ===

Id3

A14 = 13260-205934

| A5 = 200-800: YES

| A5 = 801-1400: null

| A5 = 1401-2000: null

| A5 = 2001-2600: NO

| A5 = 2601-3200: NO

| A5 = 3201-3800: null

| A5 = 3801-4400: NO

| A5 = 4400-5000: null

A14 = 205935-398608

| A3 = 500-48450: YES

| A3 = 48451-96400: YES

| A3 = 96401-144350: null

| A3 = 144351-192300: null

| A3 = 192301-240250: null

| A3 = 240251-288200: null

| A3 = 288201-336150: NO

| A3 = 336151-384100: NO

| A3 = 384101-432050: null

| A3 = 432051-480000: null

A14 = 398609-591282

| A3 = 500-48450: YES

| A3 = 48451-96400: YES

| A3 = 96401-144350: YES

| A3 = 144351-192300: YES

| A3 = 192301-240250: null

| A3 = 240251-288200: null

| A3 = 288201-336150: null
| A3 = 336151-384100: null
| A3 = 384101-432050: NO
| A3 = 432051-480000: NO
A14 = 591282-783956: YES
A14 = 783957-967630: YES
A14 = 967631-1169304: YES
A14 = 1169305-1361978: YES
A14 = 1361979-1554652: YES
A14 = 1554653-1747362: YES
A14 = 1747362-1940000: YES

El resultado de aplicar el algoritmo ID3 con la opción *Use training set* demuestra ser favorables debido a que las 60 instancias que se utilizaron fueron clasificadas de forma correcta, lo que quiere decir que el porcentaje de clasificación de las instancias es de 100.

El AD que se muestra a continuación se realizó aplicando el clasificador de tipo árbol de decisión ADTree, fue obtenido basándose en los resultados del procesamiento de la Base de Hechos. [\(Ver Anexo 2\)](#)

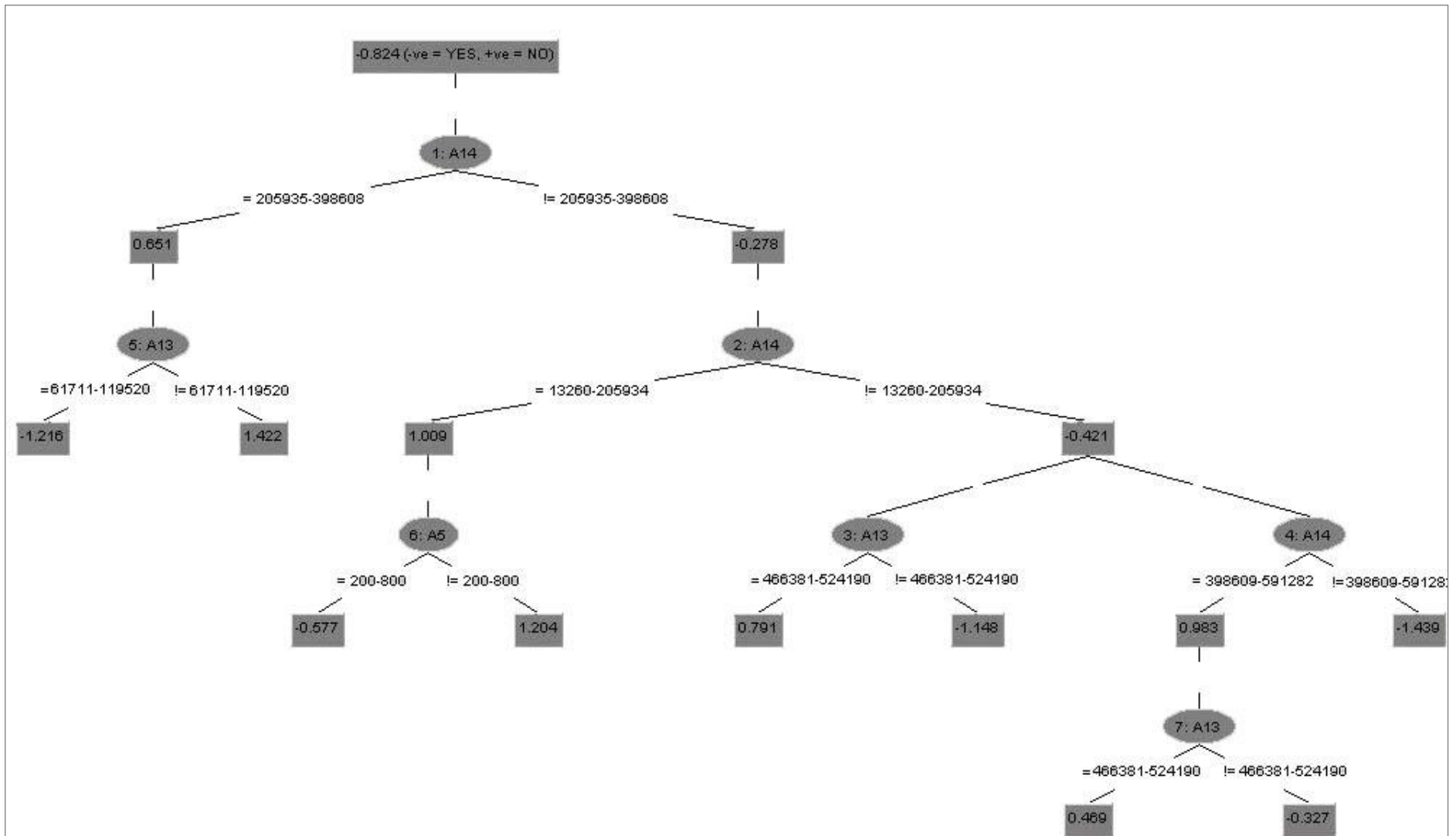


Figura 17: Árbol de Decisión con ADTree.

Fuente: Resultado de aplicar ADTree en el software WEKA.

Como se observa el AD anterior consta de ocho niveles y el atributo raíz. En él aparecen los atributos que representan la información más relevante para clasificar los proyectos como “rentables” o “no rentables”. En la muestra de entrenamiento (60 proyectos) se obtiene que 49 instancias tienen clasificaciones correctas con un porcentaje de 81.6667. De él es posible derivar las reglas de producción representadas por el camino que se utiliza para llegar a cada hoja (clase).

3.3.1 Pruebas.

Una vez obtenido el AD se realizaron dos pruebas que no cambian en nada el resultado del AD, ni la aplicación del algoritmo ID3 mediante el cual se obtuvo, sino que permite saber qué por ciento de clasificación de instancias correctas e incorrectas se logran a partir de él. Las dos pruebas que se le aplicaron al AD fueron hechas en el propio software WEKA, el cual además de permitir realizar diversos tipos de prueba, visualiza los valores que se obtienen de las mismas. Durante este proceso se utiliza un conjunto de entrenamiento que es aquel a partir del cual se obtiene el conocimiento y un conjunto de prueba que nos permite medir el alcance de ese conocimiento.

3.3.1.1 Resultados obtenidos en las pruebas.

Para la primera prueba se escogió la opción *Use training set* con la que los datos de entrada son los mismos para el conjunto de entrenamiento que de prueba, es decir, se utiliza la Base de Hechos completa para ambas operaciones. Los resultados obtenidos de esta primera prueba se muestran a continuación:

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances  60  100 %
Incorrectly Classified Instances  0  0 %
Kappa statistic      1
Mean absolute error   0.0851
Root mean squared error  0.1292
Relative absolute error  32.3748 %
Root relative squared error  36.1709 %
Total Number of Instances  60
```

Con esta prueba se obtuvo que de las 60 instancias que tiene la Base de Hechos, que cincuenta y nueve fueron correctamente clasificadas, lo que da la seguridad de que se cubre una amplia cantidad de posibilidades con este árbol.

Para la segunda prueba se escogió la opción Cross-validation donde el clasificador es evaluado por la validación cruzada, utilizando el número de pliegues que se introducen en el campo de texto Folds que para este caso fue de 9.

Seguidamente se observan los resultados:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 52 86.6667 %

Incorrectly Classified Instances 8 13.3333 %

Kappa statistic 0.4771

Mean absolute error 0.1747

Root mean squared error 0.3243

Relative absolute error 66.1895 %

Root relative squared error 90.7405 %

Total Number of Instances 60

En el caso de esta prueba, los resultados también demuestran ser favorables debido a que de 60 instancias que se utilizaron como prueba cincuenta y dos fueron clasificados de forma correcta. Lo que quiere decir que el por ciento de clasificación de instancias incorrectas es muy bajo.

Por tanto, se llegó a la conclusión de que el conocimiento que representa este árbol es fiable y puede ser utilizado como base de la aplicación a crear, porque las dos pruebas que se le aplicaron demostraron que el por ciento de clasificación de instancia es elevado, sirviendo esto como garantía de exactitud a la hora de realizar el análisis de costo de los proyectos.

3.4 Sistema de Reglas.

Tanto en el software ROSE como en el WEKA se obtuvieron buenos resultados de la calidad de la clasificación, por lo que las reglas de producción podrán ser generadas en ambos software. Se decidió trabajar con el WEKA porque en las reglas generadas se encuentran implícitamente las del software ROSE.

3.4.1 Reglas generadas por el WEKA.

Se generaron un conjunto de reglas aplicando varios clasificadores de tipo reglas. En la siguiente tabla se muestran los resultados obtenidos.

Tipo del clasificador	Calidad de la clasificación (%)	Tiempo de ejecución (s)
NNge	100	0.05
ConjunctiveRule	85	0.03
DecisionTable	91.66	0.08
JRip	88.3333	0.02
OneR	88.33	0
PART	98.3333	0.06
Ridor	85	0.02
ZeroR	85	0

Tabla 13: 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:

Regla #1: class YES IF : A1 in {4376-5500,6626-7750,7751-8875} ^ A2 in {1076-1950,4576-5450} ^ A3 in {48451-96400,192301-240250,384101-432050} ^ A4 in {501-800,801-1100,1101-1400} ^ A5 in {1401-2000} ^ A6 in {181-260,261-340} ^ A7 in {149-236,237-324,413-500} ^ A8 in {60-149,238-325} ^ A9 in {25-60,131-165,166-200} ^ A10 in {148-235,324-411,412-500} ^ A11 in {9334-17166,40666-48498} ^ A12 in {579-1081,2088-2590} ^ A13 in {61711-119520,235141-292950,466381-524190} ^ A14 in {205935-398608,783957-967630,967631-1169304} (3)

Regla #2: class YES IF : A1 in {1000-2125,2126-3250,3251-4375,4376-5500,5501-6625} ^ A2 in {200-1075,1076-1950,1951-2825,2826-3700} ^ A3 in {500-48450,48451-96400,144351-192300,240251-288200,288201-336150,336151-384100,384101-432050} ^ A4 in {200-500,501-800,801-1100,1101-1400,1401-1700,1701-2000} ^ A5 in {200-800,801-1400,1401-2000,4400-5000} ^ A6 in {100-180,261-340,341-420,421-500} ^ A7 in {60-148,149-236,325-412} ^ A8 in {60-149,150-237,238-325,326-413} ^ A9 in {25-60,61-95,96-130,131-165,166-200} ^ A10 in {60-147,148-235,236-323,324-411} ^ A11 in {1500-9333,17167-24999} ^ A12 in {75-578,1082-1584} ^ A13 in {3900-61710,61711-119520,177331-235140,235141-292950,292951-350760,350761-408570,408571-466380} ^ A14 in {13260-205934,205935-398608,591282-783956,783957-967630,967631-1169304,1169305-1361978,1361979-1554652} (10)

Regla #3: class YES IF : A1 in {1000-2125,2126-3250,3251-4375} ^ A2 in {200-1075,1076-1950} ^ A3 in {500-48450,144351-192300,192301-240250,336151-384100} ^ A4 in {200-500,501-800} ^ A5 in {2001-2600,2601-3200,3201-3800,3801-4400} ^ A6 in {100-180,181-260,421-500} ^ A7 in {60-148,237-324,413-500} ^ A8 in {60-149,238-325,326-413,414-500} ^ A9 in {25-60,61-95,96-130} ^ A10 in {60-147,324-411} ^ A11 in {1500-9333} ^ A12 in {75-578} ^ A13 in {3900-61710,177331-235140,235141-292950,350761-408570} ^ A14 in {591282-783956,783957-967630,1169305-1361978} (5)

Regla #4: class YES IF : A1 in {2126-3250,3251-4375,4376-5500,5501-6625,6626-7750,7751-8875,8876-10000} ^ A2 in {1076-1950,1951-2825,2826-3700,3701-4575,4576-5450,5451-6325,6326-7200} ^ A3 in {500-48450,48451-96400,96401-144350,144351-192300,192301-240250,240251-288200,336151-384100,384101-432050,432051-480000} ^ A4 in {200-500,501-800,801-1100,1101-1400,1401-1700,1701-2000} ^ A5 in {200-800,801-1400,2001-2600,2601-3200,3201-3800,3801-4400,4400-5000} ^ A6 in {100-180,181-260,261-340,341-420,421-500} ^ A7 in {60-148,149-236,237-324,325-412,413-500} ^ A8 in {60-149,150-237,238-325,326-413,414-500} ^ A9 in {25-60,61-95,96-130,131-165,166-200} ^ A10 in {60-147,148-235,236-323,324-411,412-500} ^ A11 in {9334-17166,32833-40665,40666-48498,48499-56331,56332-64164,64165-72000} ^ A12 in {75-578,579-1081,1585-2087,2088-2590,2591-3094,3095-3600} ^ A13 in {3900-61710,61711-119520,119521-177330,177331-235140,235141-292950,292951-350760,350761-408570,408571-466380,466381-524190,524190-582000} ^ A14 in {205935-398608,398609-591282,591282-783956,783957-967630,967631-1169304,1169305-1361978,1361979-

1554652,1554653-1747362,1747362-1940000} (30)

Regla #5: class YES IF : A1 in {2126-3250,3251-4375} ^ A2 in {200-1075,1076-1950} ^ A3 in {500-48450,144351-192300} ^ A4 in {1701-2000} ^ A5 in {2601-3200,3201-3800} ^ A6 in {181-260,341-420} ^ A7 in {413-500} ^ A8 in {60-149,150-237} ^ A9 in {25-60,96-130} ^ A10 in {60-147} ^ A11 in {1500-9333} ^ A12 in {75-578} ^ A13 in {61711-119520,119521-177330} ^ A14 in {205935-398608,398609-591282} (2)

Regla #6: class NO IF : A1 in {4376-5500} ^ A2 in {1951-2825} ^ A3 in {336151-384100} ^ A4 in {801-1100} ^ A5 in {801-1400} ^ A6 in {261-340} ^ A7 in {413-500} ^ A8 in {238-325} ^ A9 in {166-200} ^ A10 in {148-235} ^ A11 in {17167-24999} ^ A12 in {1082-1584} ^ A13 in {292951-350760} ^ A14 in {205935-398608} (1)

Regla #7: class NO IF : A1 in {7751-8875} ^ A2 in {6326-7200} ^ A3 in {384101-432050} ^ A4 in {1401-1700} ^ A5 in {1401-2000} ^ A6 in {181-260} ^ A7 in {237-324} ^ A8 in {414-500} ^ A9 in {25-60} ^ A10 in {148-235} ^ A11 in {64165-72000} ^ A12 in {3095-3600} ^ A13 in {466381-524190} ^ A14 in {398609-591282} (1)

Regla #8: class NO IF : A1 in {7751-8875} ^ A2 in {5451-6325} ^ A3 in {288201-336150} ^ A4 in {200-500} ^ A5 in {801-1400} ^ A6 in {100-180} ^ A7 in {60-148} ^ A8 in {60-149} ^ A9 in {96-130} ^ A10 in {412-500} ^ A11 in {48499-56331} ^ A12 in {2591-3094} ^ A13 in {350761-408570} ^ A14 in {205935-398608} (1)

Regla #9: class NO IF : A1 in {1000-2125,2126-3250,3251-4375,4376-5500} ^ A2 in {200-1075,1076-1950,1951-2825,2826-3700} ^ A3 in {500-48450,144351-192300,288201-336150,336151-384100,432051-480000} ^ A4 in {200-500,801-1100,1701-2000} ^ A5 in {2001-2600,2601-3200,3201-3800,3801-4400} ^ A6 in {100-180,181-260,261-340,421-500} ^ A7 in {149-236,237-324,325-412,413-500} ^ A8 in {60-149,150-237,238-325} ^ A9 in {61-95,96-130,131-165,166-200} ^ A10 in {60-147,236-323,324-411,412-500} ^ A11 in {1500-9333,17167-24999,25000-32832} ^ A12 in {75-578,1082-1584} ^ A13 in {3900-61710,177331-235140,292951-350760,350761-408570,466381-524190} ^ A14 in {13260-205934,205935-398608,398609-591282} (6)

Regla #10: class YES IF : A1 in {6626-7750} ^ A2 in {3701-4575} ^ A3 in {288201-336150} ^ A4 in {200-500} ^ A5 in {801-1400} ^ A6 in {421-500} ^ A7 in {149-236} ^ A8 in {60-149} ^ A9 in {131-165} ^ A10 in

{324-411} ^ A11 in {32833-40665} ^ A12 in {1585-2087} ^ A13 in {350761-408570} ^ A14 in {1169305-1361978} (1)

De las reglas generadas anteriormente, el número final indica la cantidad de objetos que pertenecen a ella de un total de 60 instancias.

3.5 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.
- Se clasifica como un sistema basado en reglas. Las reglas que lo soportan son las obtenidas en el AD.
- Consta de una sola clase nombrada Analizador, la misma contiene como atributos los correspondientes a las instancias de las Bases de Hecho: Materias Primas Materiales, Portadores Energéticos, Otros Gastos Servicios, Otros Gastos Comerciales, Otros Tributos, Comunicaciones, Servicios Mantenimiento Reparación, Servicios Inmobiliarios Protección, Seguros, Servicios Varios, Salarios, Impuestos contribución, Costo Producción y Precio Venta.
- Permite definir dada una entrada de datos si un proyecto es rentable o no y el margen de utilidad en %.

3.6 Interfaces de la aplicación.

La aplicación consta de tres ventanas: la principal, Analizador de costo que a partir de los datos entrados analiza si un proyecto es rentable es no y la ventana Ayuda cuya función es mostrar una síntesis de cómo funciona la aplicación. A través de las interfaces que conforman la aplicación se explicará su funcionamiento. La secuencia de imágenes que se muestra a continuación representa todo el proceso a seguir para realizar un análisis de costo.



Figura 18: Interfaz de inicio.

La interfaz de inicio es lo primero que se le muestra al usuario al abrir la aplicación, en ella se encuentra el logotipo y el nombre de la aplicación. Pinchando sobre el botón Operaciones se comienza el proceso de analizar el costo de un proyecto.

Análisis de Costos

Datos de Análisis

Gastos Estimados

	Rango (CUC)		Rango (CUC)
Materias Primas y Materiales:	5612 (1000 - 10 000)	Servicios de Mantenimiento y Reparación:	314 (60 - 500)
Portadores Energéticos:	3398 (200 - 7200)	Servicios inmobiliarios y de Protección :	366 (60 - 500)
Otros Gastos y Servicios:	421169 (500 - 480000)	Seguros:	67 (25 - 200)
Otros Gastos Comerciales:	1791 (200 - 2000)	Servicios Varios:	222 (60 - 500)
Otros Tributos:	2594 (200 - 5000)	Salarios:	37929 (1500 - 72000)
Comunicaciones:	343 (100 - 500)	Impuestos y contribución:	1896,45 (75 - 3600)

Cálculos Necesarios

Costo de Producción:	475701,45 (3900 - 582000)	Propuesta del Costo de Producción	475701,45
Precio de Venta :	1585671,50 (13260 - 1940000)	Propuesta del Precio Venta	1585671,5

El proyecto es Rentable con un margen de utilidad: 0,7

Limpiar Campos Calcular Rentabilidad Cerrar

Figura 19: Interfaz de Analizador.

En la interfaz Analizador se le solicita al usuario que entre los datos correspondientes a los análisis de costo en correspondencia al rango de valores que soporta la aplicación, donde de acuerdo con los valores que se le introdujeron a la aplicación muestra como respuesta si el proyecto es rentable o no. Si el usuario deja alguno de estos campos vacíos se muestra un mensaje indicándolo: "Debe llenar los campos vacíos."

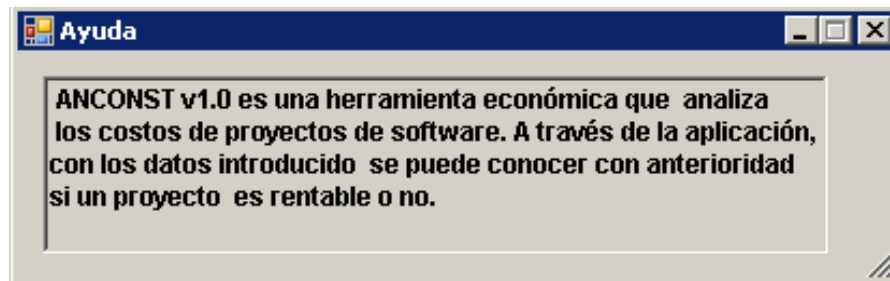


Figura 20: Interfaz de Ayuda.

La interfaz de Ayuda le muestra al usuario una breve descripción de la aplicación.

3.7 Pruebas realizadas a la aplicación.

Para verificar que la aplicación tiene un correcto funcionamiento se le aplicaron algunas pruebas. Se utilizaron juegos de datos como entrada, conociéndose la salida que se espera con ellos. A continuación se muestran las pruebas aplicadas a dos juegos de datos distintos, uno para cada uno de las posibles salidas que puede brindar la aplicación.

Juego de Datos 1

Entrada: Materias Primas Materiales 4053, Portadores Energéticos 2253, Otros Gastos Servicios 13475, Otros Gastos Comerciales 324, Otros Tributos 2516, Comunicaciones 311, Servicios Mantenimiento Reparación 341, Servicios Inmobiliarios Protección 125, Seguros 108, Servicios Varios 387, Salarios 23212, Impuestos contribución 1160.60, Costo Producción 48265.60 y Precio Venta 48145.60.

Salida Esperada: El proyecto según sus características no es rentable.

Salida Obtenida: El proyecto según sus características no es rentable.

Juego de Datos 2

Entrada: Materias Primas Materiales 5612, Portadores Energéticos 3398, Otros Gastos Servicios 421169, Otros Gastos Comerciales 1791, Otros Tributos 2594, Comunicaciones 343, Servicios Mantenimiento Reparación 314, Servicios Inmobiliarios Protección 366, Seguros 67, Servicios Varios 222, Salarios 37929, Impuestos contribución 1896.45, Costo Producción 475701.45 y Precio Venta 1585651.50.

Salida Esperada: El proyecto según sus características es rentable.

Salida Obtenida: El proyecto según sus características es rentable.

Las pruebas aplicadas con los distintos juegos de datos, se consideran satisfactorias. En ambos casos los resultados obtenidos estuvieron en relación con los esperados. Se plantea por esta razón que el sistema creado, teniendo en cuenta que los resultados que brinda vienen dados por el conocimiento que él mismo tiene almacenado es bastante confiable.

CONCLUSIONES

Durante el proceso de investigación se realizó la consulta a expertos para conocer las relaciones existentes entre las variables de análisis de costo utilizadas; de acuerdo con el análisis estadístico desarrollado en el software Statgraphics, fue aceptado el criterio de los expertos, validando la importancia del inicio de la investigación.

Se confeccionó un modelo del cual se extraen el conjunto de los atributos más importantes para el análisis de costo del proyecto seleccionado; con ellos se obtuvo una Bases de Hechos donde para su correcta creación se consultaron un grupo de especialistas.

La obtención del conocimiento se efectuó utilizando el software WEKA y los resultados obtenidos fueron sometidos a un proceso de prueba en el propio software. Del conocimiento adquirido se derivó un conjunto de reglas, utilizadas como soporte para la aplicación, por lo que la misma es considerada un sistema basado en reglas, a la cual se le aplicaron pruebas con juegos de datos obteniéndose resultados favorables.

Se desarrolló una aplicación en la que se emplea la teoría de los conjuntos aproximados para facilitar la realización del análisis de costo del proyecto productivo SCADA. A través de la aplicación creada se puede conocer con anterioridad si un proyecto es rentable o no.

RECOMENDACIONES

Profundizar en el tema de la teoría de los conjuntos aproximados ya que el mismo es poco desarrollado en la UCI.

Trabajar con datos reales para obtener mejores resultados.

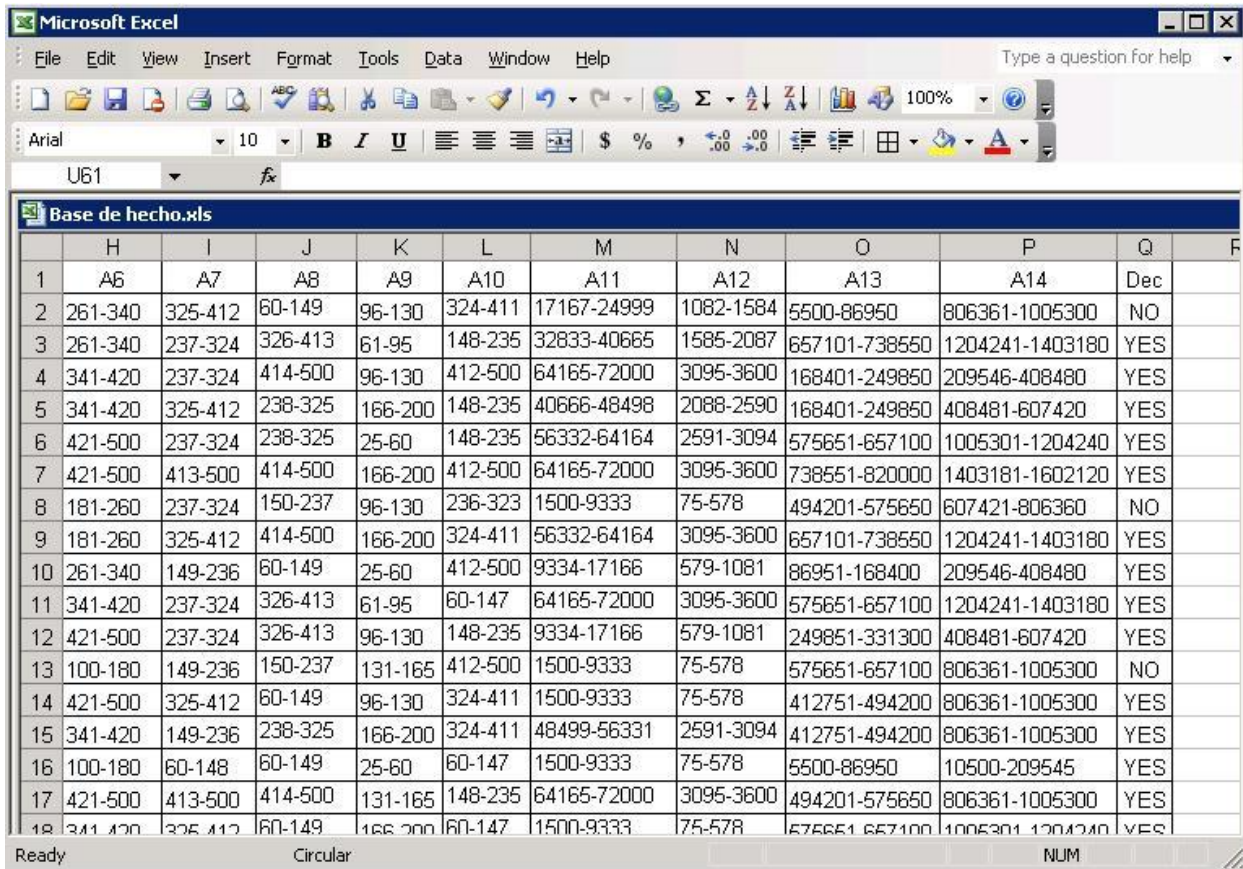
Que en próximos trabajos se utilice la rentabilidad desde diferentes puntos de vista: económicos, científicos y sociales.

Continuar trabajando como tema investigativo en la perfección del software para que sea utilizada no solo en la UCI, sino también en otras empresas que así lo requieran.

Se le propone a ALBET que siga trabajando en la mejora de esta aplicación en una versión Web para que pueda ser aplicada a todos los proyectos de la UCI.

ANEXOS

Anexo 1



The screenshot shows a Microsoft Excel window titled "Base de hecho.xls". The spreadsheet contains a table with 11 columns (H through Q) and 18 rows. The data is as follows:

	H	I	J	K	L	M	N	O	P	Q	F
1	A6	A7	A8	A9	A10	A11	A12	A13	A14	Dec	
2	261-340	325-412	60-149	96-130	324-411	17167-24999	1082-1584	5500-86950	806361-1005300	NO	
3	261-340	237-324	326-413	61-95	148-235	32833-40665	1585-2087	657101-738550	1204241-1403180	YES	
4	341-420	237-324	414-500	96-130	412-500	64165-72000	3095-3600	168401-249850	209546-408480	YES	
5	341-420	325-412	238-325	166-200	148-235	40666-48498	2088-2590	168401-249850	408481-607420	YES	
6	421-500	237-324	238-325	25-60	148-235	56332-64164	2591-3094	575651-657100	1005301-1204240	YES	
7	421-500	413-500	414-500	166-200	412-500	64165-72000	3095-3600	738551-820000	1403181-1602120	YES	
8	181-260	237-324	150-237	96-130	236-323	1500-9333	75-578	494201-575650	607421-806360	NO	
9	181-260	325-412	414-500	166-200	324-411	56332-64164	3095-3600	657101-738550	1204241-1403180	YES	
10	261-340	149-236	60-149	25-60	412-500	9334-17166	579-1081	86951-168400	209546-408480	YES	
11	341-420	237-324	326-413	61-95	60-147	64165-72000	3095-3600	575651-657100	1204241-1403180	YES	
12	421-500	237-324	326-413	96-130	148-235	9334-17166	579-1081	249851-331300	408481-607420	YES	
13	100-180	149-236	150-237	131-165	412-500	1500-9333	75-578	575651-657100	806361-1005300	NO	
14	421-500	325-412	60-149	96-130	324-411	1500-9333	75-578	412751-494200	806361-1005300	YES	
15	341-420	149-236	238-325	166-200	324-411	48499-56331	2591-3094	412751-494200	806361-1005300	YES	
16	100-180	60-148	60-149	25-60	60-147	1500-9333	75-578	5500-86950	10500-209545	YES	
17	421-500	413-500	414-500	131-165	148-235	64165-72000	3095-3600	494201-575650	806361-1005300	YES	
18	341-420	325-412	60-149	166-200	60-147	1500-9333	75-578	575651-657100	1005301-1204240	YES	

Figura 21: Base de Hechos en Excel.

Anexo 2

The screenshot shows the Weka 3.5.8 Explorer window. The 'Current relation' is 'Costo' with 60 instances and 15 attributes. The 'Selected attribute' is 'A1', which is a nominal attribute with 8 distinct values and 0 missing values. The 'Attributes' list shows 14 attributes (A1 to A14) with checkboxes. The 'Visualize All' button is active, and a bar chart is displayed below the attribute details. The bar chart shows the distribution of the selected attribute 'A1' across the instances. The bars are blue with red segments on top, representing the count of instances for each attribute value.

Label	Count
1000-2125	12
2126-3250	5
3251-4375	8
4376-5500	7
5501-6625	3
6626-7750	4
7751-8875	11
8876-10000	10

Figura 22: Procesamiento de la Base de Hechos en WEKA.

Anexo 3

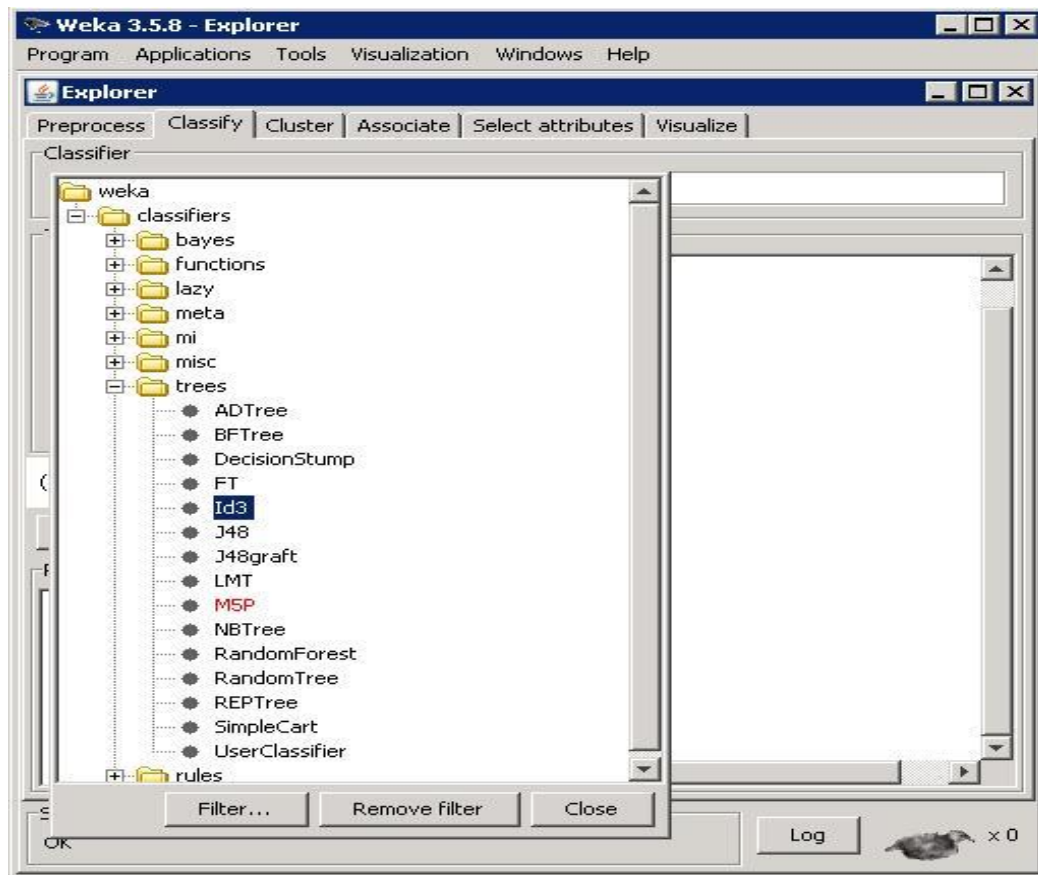


Figura 23: Clasificación en WEKA.

Anexo 4

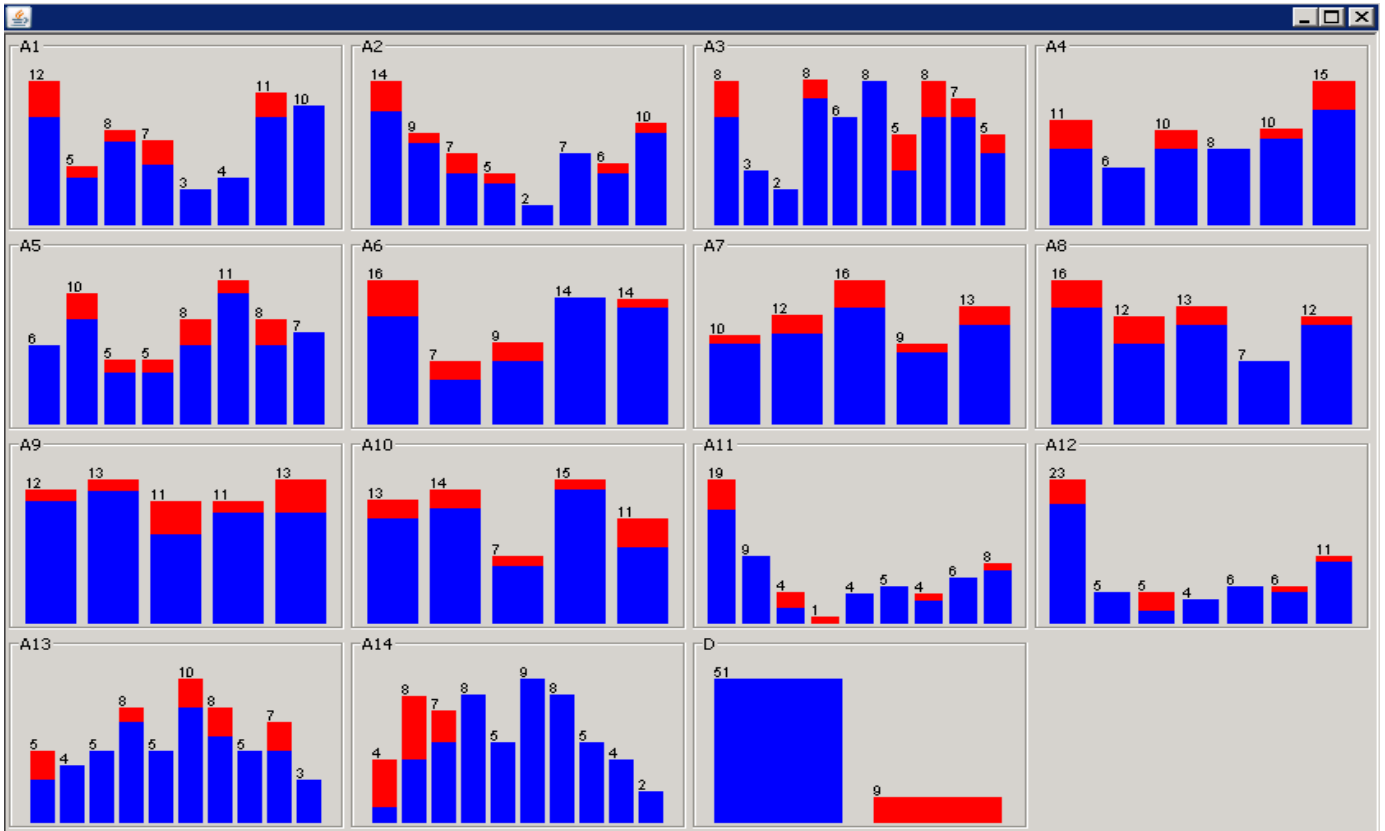


Figura 24: Representación del sistema de decisión en WEKA.

CITAS BIBLIOGRÁFICAS

1. **Rodríguez, Manuel.** *La inteligencia artificial como herramienta de modelación en la predicción de la insolvencia empresarial* [Disponible] 2009 [Citado: junio 4, 009.]
http://www.navactiva.com/web/es/descargas/pdf/acyf/inteligencia_artificial_2.pdf
2. **Álvarez Teruelo, Sandra; Fernández Pacheco, Iván Alejandro.** *Análisis de Predicción de Terremotos.* Universidad Carlos III Madrid. [Citado: marzo 25, 2010].
<http://www.it.uc3m.es/jvillena/irc/practicas/07-08/PrediccionTerremotos.pdf>
3. **Edgar.** *Clasificación de los costos de operaciones.* [Citado: febrero 8, 010]
http://mexinetica.com/~lanjoe9/costos05-2/Equipo07-Tema05Clasificacion_de_los_costos_de_operacion..ppt
4. *Costos de producción.* [Citado: febrero 8, 010]
<http://www.fao.org/DOCREP/003/v8490s/v8490s06.htm>
5. **Segovia Vargas, María Jesús.** *Predicción de crisis empresariales en seguros no vida mediante la metodología rough set.* [Disponible] 2003 [Citado: noviembre 6, 009.]
<http://eprints.ucm.es/tesis/cee/ucm-t26780.pdf>
6. **Benítez, Miguel Ángel.** *Contabilidad y Finanzas para la Formación Económica de los Cuadros de Dirección.* – Cuba, 1997. p. 150.
7. *Tablas de decisión.*[Citado: febrero 10, 010]
http://paginasweb.univalle.edu.co/~cimpa/presentaciones/descargar_archivo.php?id=90
8. **Sosa Sierra, María del Carmen.** *Importancia de la inteligencia artificial en la gerencia empresarial.* Universidad del Norte. Disponible [2007]. [Citado: febrero 10, 010]
<http://www.monografias.com/trabajos28/inteligencia-artificial-gestion-empresarial/inteligencia-artificial-gestion-empresarial.shtml>

9. **Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L.** *Predicción de insolvencias con el método Rough Set.* Universidad Complutense de Madrid, Campus de Somosaguas. [Citado: febrero 10, 010] <http://eprints.ucm.es/6802/1/0303.pdf>
10. *Selección de rasgos relevantes para el análisis de datos a través de los Conjuntos Aproximados.* [Citado: febrero 15, 010]
<http://www.bibliociencias.cu/gsd/collect/eventos/import/Selecci%F3n%20de%20rasgos%20relevantes.doc>
11. **Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L.** *La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multiatributo.* [Citado: abril 7, 010] <http://www.uv.es/asepuma/XI/19.pdf>
12. **Diccionario de economía y finanzas.** [Citado: febrero 8, 010]
<http://www.eumed.net/cursecon/dic/c13.htm>
13. *Rentabilidad del producto* [citado: febrero 16, 2010] <http://ocw.uc3m.es/economia-financiera-y-contabilidad/economia-de-la-empresa/material-de-clase-1/Rentabilidad.pdf>
14. **Aristizabal A., Hernán A.; Galvis C, Alberto.** *Análisis de costo en sistemas de potabilización de agua.* Universidad del Valle (Colombia). [Citado: enero 21, 2009.]
<http://www.cepis.org.pe/bvsacd/congreso/anacos.pdf>
15. **Quintero Álvarez, Liliana Esther.** *SISTEMAS DE COSTOS.* [Citado: febrero 8, 010.]
<http://www.uaim.edu.mx/web-carreras/carreras/contaduria/SISTEMAS%20DE%20COSTOS.pdf>
16. **Morate, Diego García.** *Manual de WEKA.* : [Disponible] 2005 [Citado: febrero 16, 2010.]
<http://metaemotion.com/diego.garcia.morate>

17. **Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L.** *La metodología Rough Set frente al análisis discriminante en la predicción de insolvencia en empresas aseguradoras.* [Citado: abril 7, 2010.] <http://www.actuarios.org/espa/anales/2003/Segovia%20Vargas2003.pdf>
18. *El método Delphi* [Citado: febrero 16, 2010.] <http://www.gtic.ssr.upm.es/encuestas/delphi.htm>
19. *Curso básico de Statgraphics* [Citado: febrero 17, 2010.] <http://www.sisoft.ucm.es/Manuales/Statgraphics7.pdf>
20. *Presupuestos de ingresos, costos y gastos* [Citado: febrero 16, 2010.] <http://www.mailxmail.com/curso-proyectos-inversion/presupuestos-ingresos-costos-gastos>
21. **Álvarez Peña, Migdalia; Cosano Delgado, Saskia; Chacón Pupo, Yamilet.** *Análisis Económico Financiero en la Sucursal Cimex de Las Tunas.* 2005 [Citado: abril 5, 2010.] <http://www.gestiopolis.com/finanzas-contaduria/analisis-economico-financiero-en-tunas.htm>
22. **Corazón Del Sol, Eliades.** *Análisis Económico Financiero en la Empresa de Materiales de construcción de Las Tunas.* 2005 [Citado: abril 5, 2010.] <http://www.gestiopolis.com/finanzas-contaduria/analisis-economico-financiero-en-una-empresa-de-materiales.htm>
23. *Diplomado de Formación de Promotores, Consultores y Técnicos en Microempresa.* [Citado: marzo 23, 2010]. <http://www.infomipyme.com/Docs/GT/Tecnicos/pdf/0045.pdf> - Guatemala
24. **Esterkin, José.** *La estimación del costo de un proyecto.* [Disponible] febrero 18, 2008 [Citado: febrero 22, 2010.] <http://iaap.wordpress.com/2008/02/18/la-estimacion-del-costode-un-proyecto/>
25. *Economía de la empresa. Análisis de las decisiones empresariales.* [Citado: febrero 16, 2010.] http://bibliodoc.uci.cu/pdf/eco_empresa1.pdf
26. *Diagnóstico financiero.* [Citado: febrero 13, 2010.] http://www.angelfire.com/ex2/john_rosso/Posgrados/A_FIN.pdf

27. **Costos.** [Citado: febrero 21, 2010.]
<http://www.infomipyme.com/Docs/GT/Offline/Empresarios/costos.htm>
28. **Díaz Arévalo, José Luis; Pérez García, Rafael.** *Estado del arte en la utilización de técnicas avanzadas para la búsqueda de información no trivial a partir de datos en los sistemas de abastecimiento de agua potable.* Universidad Politécnica de Valencia.
29. *Algunas definiciones de Inteligencia Artificial* [Citado: febrero 16, 2010.]
<http://www.infor.uva.es/~calonso/IAI/Tema1-introduccionIAI/DefinicionesIA.pdf>
30. **Bello Pérez, Rafael.** *La teoría de los conjuntos aproximados en el contexto de la inteligencia artificial.* Universidad Central de las Villas. Cuba. Disponible [2008]. [Citado: noviembre 10, 2010]
http://inter-nos.uci.cu/Teleclases/Teleclases.asp?id_as=33
31. **Álvarez Laverde, Héctor René.** *El modelo Rough Sets y su aplicación en la Ingeniería Kansei.* Barcelona. 2009
32. **Serrano González, Miguel García.** *Evolución de proyecciones lineales para aprendizaje automático.* Universidad Carlos III de Madrid. Disponible [2009]. [Citado: marzo 20, 2010]. http://e-archivo.uc3m.es/bitstream/10016/6460/1/PFC_Miguel_Garcia-Serrano_Gonzalez.pdf
33. **Vizcaino Garzon, Paula Andrea.** *Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (Waikato Environment for Knowledge Analysis).* Facultad de ingeniería de sistema de Bogotá. Disponible [2008]. [Citado: marzo 20, 2010].
http://www.fukl.edu.co/images/stories/suma_digital_sistemas/2009_01/final_paula_andrea.pdf
34. **Macías Rodríguez, Miguel.** *Técnicas de minería de datos para la retención de clientes en el sector de asegurador.* Disponible [2008]. [Citado: marzo 20, 2010].
<http://www.cnsf.gob.mx/Eventos/Premios/2008%20Seguros/ANIVDELAREV.pdf>

35. **Obregón Pizarro, Víctor Eleno.** *Análisis de Sistema de Costos Económicos en los Astilleros Menores y Propuestas de Nuevas Innovaciones Lima – Callao.* Universidad Nacional de Ingeniería (Lima – Perú). [Citado: enero 21, 2009.]
http://www.ipen.org.br/downloads/XIX/CT3_CONSTRUCCIONES_NAVALES/Victor_Obregon.pdf
36. **Trujillo, Alberto; Cárdenas, Carmen; Saules, Luisa; Fernández, Francisca; Zapata, Yralih; Zapata, Yralih.** *Determinación y análisis de costos por el funcionamiento, operación y mantenimiento del sistema piloto de lagunas de estabilización de la Universidad del Zulia.* Universidad del Zulia. (Maracaibo, Venezuela). [Citado: enero 21, 2009.]
<http://www.bvsde.paho.org/bvsaidis/mexico26/ii-134.pdf>
37. **Cabrera Jiménez, Yenisbel; Díaz Rodríguez, Beatriz; Amat Álvarez, Erizbel; Cabrera Pérez, Nallivis; González Cabrera, Yunet; Jiménez Padilla, Tamara; Lara Zayas, María Luisa.** *El Análisis Económico Financiero como Herramienta de Dirección. Generalidades.* [Citado: abril 5, 2010.] <http://www.gestiopolis.com/recursos6/Docs/Fin/analisis-economico-financiero.htm>
38. **Caballero, Yailé; Bello, Rafael; Alvarez, Delia; García, María M.; Baltá, Analay.** *Un nuevo algoritmo de selección de rasgos basado en la Teoría de los Conjuntos Aproximados.* [Disponible] 2007 [Citado: noviembre 17, 2009.] <http://redalyc.uaemex.mx/redalyc/pdf/430/43004111.pdf>

BIBLIOGRAFÍA

Caballero, Yailé; Bello, Rafael; Alvarez, Delia; García, Maria M.; Baltá, Analay. *Un nuevo algoritmo de selección de rasgos basado en la Teoría de los Conjuntos Aproximados.* [Disponible] 2007 [Citado: noviembre 17, 009.] <http://redalyc.uaemex.mx/redalyc/pdf/430/43004111.pdf>

Segovia Vargas, María Jesús. *Predicción de crisis empresariales en seguros no vida mediante la metodología rough set.* [Disponible] 2003 [Citado: noviembre 6, 009.] <http://eprints.ucm.es/tesis/cee/ucm-t26780.pdf>

Álvarez Teruelo, Sandra; Fernández Pacheco, Iván Alejandro. *Análisis de Predicción de Terremotos.* Universidad Carlos III Madrid. [Citado: marzo 25, 2010]. <http://www.it.uc3m.es/jvillena/irc/practicas/07-08/PrediccionTerremotos.pdf>

Diplomado de Formación de Promotores, Consultores y Técnicos en Microempresa. [Citado: marzo 23, 2010]. <http://www.infomipyme.com/Docs/GT/Tecnicos/pdf/0045.pdf> - Guatemala

Sosa Sierra, María Del Carmen. *Inteligencia artificial en la gestión financiera empresarial.* Agosto 2007.

Quintero Álvarez, Liliana Esther. *SISTEMAS DE COSTOS.* [Citado: febrero 8, 010.] <http://www.uaim.edu.mx/web-carreras/carreras/contaduria/SISTEMAS%20DE%20COSTOS.pdf>

Edgar. *Clasificación de los costos de operaciones.* [Citado: febrero 8, 010] http://mexinetica.com/~lanjoe9/costos05-2/Equipo07-Tema05-Clasificacion_de_los_costos_de_operacion..ppt

Costos de producción. [Citado: febrero 8, 010] <http://www.fao.org/DOCREP/003/v8490s/v8490s06.htm>

Kirkby, Richard; Frank, Eibe; Reutemann, Peter. *WEKA Explorer User Guide for Version 3-5-8.* [Disponible] 2008 [Citado: febrero 16, 2010.] <http://www.cs.waikato.ac.nz/~ml/WEKA/>

Statgraphics [Citado: febrero 16, 2010.] <http://www.softwarecientifico.com/paginas/statgrap1.htm>

Curso básico de Statgraphics [Citado: febrero 17, 2010.]
<http://www.sisoft.ucm.es/Manuales/Statgraphics7.pdf>

El método Delphi [Citado: febrero 16, 2010.] <http://www.gtlic.ssr.upm.es/encuestas/delphi.htm>

Rentabilidad del producto [citado: febrero 16, 2010] <http://ocw.uc3m.es/economia-financiera-y-contabilidad/economia-de-la-empresa/material-de-clase-1/Rentabilidad.pdf>

Restrepo López, Mauricio; Andrés López, Héctor. *Uso de la metodología Rough Sets como soporte a la toma de decisiones* [Citado febrero 17, 2010].

http://paginasweb.univalle.edu.co/~cimpa/presentaciones/descargar_archivo.php?id=90

Segovia Vargas, M.J.; Gil Fana, J.A.; Heras Martínez, A.; Vilar Zanón, J.L. *Predicción de insolvencias con el método Rough Set* [Citado febrero 17, 2010]. <http://eprints.ucm.es/6802/1/0303.pdf>

Ramentol Martínez, Enislay; Herrera, Francisco; Bello Pérez, Rafael; Caballero Mota, Yailé; Sánchez López, Yanet. *Edición de Conjuntos de Entrenamiento no Balanceados, haciendo uso de Operadores Genéticos y la Teoría de los Conjuntos Aproximados.* [Citado febrero 19, 2010]

http://sci2s.ugr.es/publications/ficheros/maeb2009_submission_42.pdf

Cárdenas, Beitmantt; Caballero, Yailé; Bello, Rafael. *La Teoría de los Conjuntos Aproximados y las Técnicas de Bootstrap para la Edición de Conjuntos de Entrenamiento. Su Aplicación en el Pronóstico Meteorológico.* [Citado febrero 19, 2010]

<http://pisis.unalmed.edu.co/avances/archivos/ediciones/Edicion%20Avances%202007%203/18.pdf>

Calzadilla Fernández de Castro, Omar; Jiménez López, Giset; González Delgado, Blanca Estela; Ávila Pérez, Jenny. *Aplicación de la minería de datos al Sistema Cubano de Farmacovigilancia.* [Citado febrero 20, 2010]

<http://scielo.sld.cu/pdf/far/v41n3/far03307.pdf>

ROSE2. [Citado: febrero 19, 2010.] <http://idss.cs.put.poznan.pl/site/rose.html>

Esterkin, Jose. *La estimación del costo de un proyecto.* [Disponible] febrero 18, 2008 [Citado: febrero 22, 2010.] <http://iaap.wordpress.com/2008/02/18/la-estimacion-del-costode-un-proyecto/>

Cabrera Jiménez, Yenisbel; Díaz Rodríguez, Beatriz; Amat Álvarez, Erizbel; Cabrera Pérez, Nallivis; González Cabrera, Yunet; Jiménez Padilla, Tamara; Lara Zayas, María Luisa. *El Análisis Económico Financiero como Herramienta de Dirección. Generalidades.* [Citado: abril 5, 2010.] <http://www.gestiopolis.com/recursos6/Docs/Fin/analisis-economico-financiero.htm>

Álvarez Peña, Migdalia; Cosano Delgado, Saskia; Chacón Pupo, Yamilet. *Análisis Económico Financiero en la Sucursal Cimex de Las Tunas.* 2005 [Citado: abril 5, 2010.] <http://www.gestiopolis.com/finanzas-contaduria/analisis-economico-financiero-en-tunas.htm>

Corazón Del Sol, Eliades. *Análisis Económico Financiero en la Empresa de Materiales de construcción de Las Tunas.* 2005 [Citado: abril 5, 2010.] <http://www.gestiopolis.com/finanzas-contaduria/analisis-economico-financiero-en-una-empresa-de-materiales.htm>

Presupuestos de ingresos, costos y gastos [Citado: febrero 16, 2010.] <http://www.mailxmail.com/curso-proyectos-inversion/presupuestos-ingresos-costos-gastos>

Economía de la empresa. Análisis de las decisiones empresariales. [Citado: febrero 16, 2010.] http://bibliodoc.uci.cu/pdf/eco_empresa1.pdf

Diagnóstico financiero. [Citado: febrero 13, 2010.] http://www.angelfire.com/ex2/john_rosso/Posgrados/A_FIN.pdf

Costos. [Citado: febrero 21, 2010.] <http://www.infomipyme.com/Docs/GT/Offline/Empresarios/costos.htm>

Díaz Arévalo, José Luis; Pérez García, Rafael. *Estado del arte en la utilización de técnicas avanzadas para la búsqueda de información no trivial a partir de datos en los sistemas de abastecimiento de agua potable.* Universidad Politécnica de Valencia.

Román Carrillo, Javier; Virseda Benito, Fernando. *Minería de datos y aplicaciones.* Universidad Carlos III.

Serrano González, Miguel García. *Evolución de proyecciones lineales para aprendizaje automático.* Universidad Carlos III de Madrid. Disponible [2009]. [Citado: marzo 20, 2010]. http://e-archivo.uc3m.es/bitstream/10016/6460/1/PFC_Miguel_Garcia-Serrano_Gonzalez.pdf

Vizcaino Garzon, Paula Andrea. *Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (Waikato Environment for Knowledge Analysis).* Facultad de ingeniería de sistema de Bogotá. Disponible [2008]. [Citado: marzo 20, 2010]. http://www.fukl.edu.co/images/stories/suma_digital_sistemas/2009_01/final_paula_andrea.pdf

Macías Rodríguez, Miguel. *Técnicas de minería de datos para la retención de clientes en el sector de asegurador.* Disponible [2008]. [Citado: marzo 20, 2010]. <http://www.cnsf.gob.mx/Eventos/Premios/2008%20Seguros/ANIVDELAREV.pdf>

Díaz Martínez, Zuleyka; Fernández Menéndez, José; Segovia Vargas, M^a Jesús. *Sistema de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradas.* Universidad Complutense de Madrid. [Citado: marzo 20, 2010]. http://www.uv.es/asepuma/XII/comunica/diaz_fernandez_segovia.pdf

Sosa Sierra, María del Carmen. *Introducción a las técnicas de inteligencia artificial aplicadas a la gestión financiera empresarial.* Universidad del Norte, Colombia. [Citado: marzo 20, 2010]. <http://www.eumed.net/ce/2006/mcss.htm>

Benítez, Miguel Ángel. *Contabilidad y Finanzas para la Formación Económica de los Cuadros de Dirección.* – Cuba, 1997. p. 150.

Rodríguez, Manuel. *La inteligencia artificial como herramienta de modelación en la predicción de la insolvencia empresarial* [Disponible] 2009 [Citado: junio 4, 2009.]

http://www.navactiva.com/web/es/descargas/pdf/acyf/inteligencia_artificial_2.pdf

Tablas de decisión.[Citado: febrero 10, 2010]

http://paginasweb.univalle.edu.co/~cimpa/presentaciones/descargar_archivo.php?id=90

Sosa Sierra, María del Carmen. *Importancia de la inteligencia artificial en la gerencia empresarial.*

Universidad del Norte. Disponible [2007]. [Citado: febrero 10, 2010]

<http://www.monografias.com/trabajos28/inteligencia-artificial-gestion-empresarial/inteligencia-artificial-gestion-empresarial.shtml>

Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L. *Predicción de insolvencias con el método Rough Set.* Universidad Complutense de Madrid, Campus de Somosaguas. [Citado: febrero 10, 2010] <http://eprints.ucm.es/6802/1/0303.pdf>

Selección de rasgos relevantes para el análisis de datos a través de los Conjuntos Aproximados. [Citado: febrero 15, 2010]

<http://www.bibliociencias.cu/gsd/collect/eventos/import/Selecci%F3n%20de%20rasgos%20relevantes.doc>

Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L. *La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multiatributo.* [Citado: abril 7, 2010]

<http://www.uv.es/asepuma/XI/19.pdf>

Diccionario de economía y finanzas. [Citado: febrero 8, 2010] <http://www.eumed.net/cursecon/dic/c13.htm>

Morate, Diego García. *Manual de WEKA.* : [Disponible] 2005 [Citado: febrero 16, 2010.]

<http://metaemotion.com/diego.garcia.morate>

Segovia Vargas, M.J; Gil Fana, J.A; Heras Martínez, A; Vilar Zanón, J.L. *La metodología Rough Set frente al análisis discriminante en la predicción de insolvencia en empresas aseguradoras.* [Citado: abril 7, 2010.] <http://www.actuarios.org/espa/anales/2003/Segovia%20Vargas2003.pdf>

Curso básico de Statgraphics [Citado: febrero 17, 2010.]
<http://www.sisoft.ucm.es/Manuales/Statgraphics7.pdf>

Algunas definiciones de Inteligencia Artificial [Citado: febrero 16, 2010.]
<http://www.infor.uva.es/~calonso/IAI/Tema1-introduccionIAI/DefinicionesIA.pdf>

Bello Pérez, Rafael. *La teoría de los conjuntos aproximados en el contexto de la inteligencia artificial.* Universidad Central de las Villas. Cuba. Disponible [2008]. [Citado: noviembre 10, 2010] http://internos.uci.cu/Teleclases/Teleclases.asp?id_as=33

Álvarez Laverde, Héctor René. *El modelo Rough Sets y su aplicación en la Ingeniería Kansei.* Barcelona. 2009.

GLOSARIO DE TÉRMINOS

IA: Inteligencia Artificial.

RST: Rough Sets Theory (Teoría de los Conjuntos Aproximados).

AD: Árbol de decisión.

DCE: Departamento de Ciencias Empresariales.

ALBET. Alternativa Bolivariana para la Exportación de las Américas

UCI: Universidad de las Ciencias Informáticas.

BH: Base de Hechos.

Gastos: Los gastos expresan el monto total, en términos monetarios, de los recursos materiales y laborales utilizados en el conjunto de la actividad empresarial.

Elementos del Gasto: Permite la cuantificación de los recursos materiales, laborales y monetarios en los cuales se expresan los gastos de trabajo en el conjunto de la actividad empresarial.

Costos Estimados: Son costos que se calculan sobre bases de conocimiento y experiencia, antes de iniciarse el proceso productivo. Mediante el costo estimado se obtienen aproximaciones de lo que se debe gastar.

Toma de decisiones: Es el proceso continuo, como una serie de cambios consecutivos a partir de unos flujos de información que se experimentan en el seno del sistema.

Sistemas de Costo: Es un conjunto de métodos, normas y procedimientos que rigen la planificación, determinación y análisis del costo, así como el proceso del registro de los gastos de una o varias actividades productivas en una empresa.

Gestión de costo de un proyecto de software: Tiene como objetivo que el proyecto se complete con el presupuesto inicialmente aprobado, está relacionado tanto a los recursos materiales necesarios para el proyecto, como a los recursos humanos (horas de trabajo).

Costo de Producción: Expresa la magnitud de los recursos materiales, laborales y monetarios necesarios para alcanzar un cierto volumen de producción con una determinada calidad.

Precio de Venta: Es el valor de los productos o servicios que se venden a los clientes.

Rentabilidad: Mide el éxito o fracaso que tiene un producto.