



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE CIENCIAS DE LA INGENIERÍA

CARRERA INGENIERÍA EN SISTEMAS

Tema de la Tesis

“Representación de conocimiento a través de sistemas difusos para evaluar los estilos de enseñanza aprendizaje entre estudiantes y docentes de la UTEQ”

Previo a la obtención del título de:

Ingeniera en Sistemas

Autora:

Lely Johanna Olvera Toala

Director de Tesis

PhD.Msc. Amilkar Puris Cáceres

Quevedo – Ecuador.

Año

2015

CÓDIGO DUBLÍN

1.	Título	M	“REPRESENTACIÓN DE CONOCIMIENTO A TRAVÉS DE SISTEMAS DIFUSOS PARA EVALUAR LOS ESTILOS DE ENSEÑANZA APRENDIZAJE ENTRE ESTUDIANTES Y DOCENTES DE LA UTEQ”
2.	Creador	M	Lely Johanna Olvera Toala
3.	Materia	M	Facultad de Ciencias de la Ingeniería Análisis Inteligente de datos: Minería de datos
4.	Descripción	M	La presente investigación se realizó en la Universidad Técnica Estatal de Quevedo. El objetivo principal es representar conocimiento oculto a través de sistemas difusos para evaluar los estilos de enseñanza/ aprendizaje de la UTEQ.
5.	Editor	M	Facultad de Ciencias de la Ingeniería: Carrera de Ingeniería en Sistemas
6.	Colaboradores	0	PhD. Amilkar Puris Ing. Jorge Guanín
7.	Fecha	M	15-05-2015
8.	Tipo	M	Proyecto de Tesis
9,	Formato	M	Ms Word, Pdf
10	Identificador	M	Ninguna
11	Fuente	M	Minería de datos
12	Lenguaje	M	Español
13	Relación	0	Ninguno
14	Cobertura	0	Ninguno
15	Derechos	M	Ninguno
16	Audiencia	0	Proyecto de Investigación / Research Projec

TRIBUNAL DE TESIS



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD CIENCIAS DE LA INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS

Presentado al Consejo Directivo como requisito previo a la obtención del título
de INGENIERA EN SISTEMAS

Aprobado:

ING. CARLOS MARQUEZ DE LA PLATA
PRESIDENTE TRIBUNAL DE TESIS

ING. WASHINGTON CHIRIBOGA
MIEMBRO TRIBUNAL DE TESIS

ING. ANDREA ZUÑIGA PAREDES
MIEMBRO TRIBUNAL DE TESIS

QUEVEDO – LOS RIOS – ECUADOR

Año 2015

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **LELY JOHANNA OLVERA TOALA**, declaro que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Lely Johanna Olvera Toala

CERTIFICACIÓN DEL DIRECTOR DE TESIS

El suscrito, PhD.Msc. Amilkar Puris Cáceres, Docente de la Universidad Técnica Estatal de Quevedo, certifica que la Egresada LELY JOHANNA OLVERA TOALA, realizó la tesis de grado previo a la obtención del título de Ingeniera en Sistemas titulada **“Representación de conocimiento a través de sistemas difusos para evaluar los estilos de enseñanza aprendizaje entre estudiantes y docentes de la UTEQ”**, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

PhD. Msc. Amilkar Puris Cáceres

DEDICATORIA

A Dios por permitirme llegar a este momento tan especial en mi vida, por los triunfos y los momentos difíciles que me ha enseñado a valorarlo cada día más.

A mis padres Sr. Víctor Olvera y Sra. Lely Toala de Olvera por apoyarme de forma incondicional, quienes día a día me impulsaron a cumplir con todos mis sueños.

A mis hermanos Daniel y Josué por brindarme su compañía, cariño y por contagiarme la alegría de vivir.

A mi abuelo Sr. Gadulfo Toala por cada día consentirme y llenarme de amor

A mi enamorado Mario Pico por brindarme su apoyo incondicional

A Iván y Pablo, por brindarme su amistad durante este largo caminar.

AGRADECIMIENTO

Agradezco a Dios y a cada una de las personas que forman parte de mi vida.

A la Universidad Técnica Estatal de Quevedo por abrir sus puertas para brindarme muchos conocimientos de manera especial en la Facultad Ciencias de la Ingeniería.

Al Ing. Jorge Guanín y al PhD Amilkar Puris, porque con su contribución ha sido posible culminar la presente propuesta investigativa

A todos los docentes que me impartieron clases durante esta etapa de mi vida.

A toda mi familia y a cada una de las personas que siempre confiaron en mí

A todos ustedes muchas gracias

ÍNDICE

CÓDIGO DUBLÍN	II
TRIBUNAL DE TESIS.....	III
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	IV
CERTIFICACIÓN DEL DIRECTOR DE TESIS	V
DEDICATORIA	VI
AGRADECIMIENTO	VII
ÍNDICE DE TABLAS	XIV
RESUMEN EJECUTIVO	XVIII
ABSTRACT	XIX
CAPÍTULO I: MARCO CONTEXTUAL DE LA INVESTIGACIÓN	1
1.1. INTRODUCCIÓN	2
1.2. JUSTIFICACIÓN	5
1.3. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA	6
1.3.1. Análisis del problema	6
1.3.2. Formulación.....	6
1.3.3. Sistematización	7
1.4. OBJETIVOS	8
1.4.1. Objetivo General.....	8
1.4.2. Objetivos Específicos	8
1.5. HIPÓTESIS	9

1.5.1. Planteamiento.....	9
1.5.2. Matriz de Conceptualización.....	9
CAPÍTULO II: MARCO TEÓRICO DE LA INVESTIGACIÓN	11
2.1. FUNDAMENTACIÓN CONCEPTUAL.....	12
2.1.1. Extracción de conocimiento.....	12
2.1.2. Estilo de aprendizaje	12
2.1.3. Estrategias de Aprendizaje.....	12
2.1.4. Estilo de enseñanza	14
2.1.5. Estilo de enseñanza modelo VARK.....	14
2.1.6. Estilo de aprendizaje VARK.....	16
2.1.7. Algoritmo	17
2.1.8. Sistema Difuso	17
2.1.9. Lógica Difusa.....	18
2.1.10. Variables Difusas.....	19
2.1.11. Preprocesamiento de Datos	19
2.1.12. Discretización de características	20
2.1.13. KDD: Proceso de Extracción de conocimiento	21
2.1.14. Extracción de conocimiento en grandes volúmenes de datos	22
2.1.15. Las fases del proceso de extracción del conocimiento.....	23
2.1.16. Operaciones de preparación de datos.....	26
2.1.17. Reducción de dimensionalidad.....	27

2.1.18. Reducción del número de atributos	27
2.1.19. Reducción del número de casos	27
2.1.20. Tratamiento de la falta de datos	28
2.1.21. Evaluación de modelos clasificatorios	28
2.1.22. Software Keel	29
2.1.23. Weka	31
2.2. MARCO REFERENCIAL.....	32
2.2.1. Efecto de minería de datos educacional para mejorar el de aprendizaje mediante el sistema de aprendizaje electrónico.....	32
2.2.2. Minería de datos para guiar a los estudiantes a través del proceso de inscripción basada en el rendimiento académico	33
2.2.3. Análisis de calificación de estudiantes basados en conjuntos aproximados.....	33
2.2.4. Aplicación de minería por medio de reglas de asociación en el sistema de evaluación docente	33
2.2.5. Personalización de los procesos de aprendizaje de la minería de datos.....	34
2.2.6. Aplicación de minería de datos educativa para predecir las tendencias académicas y los patrones en las base de datos	34
2.2.7. Extracción y análisis del desempeño docente de la gestión de retroalimentación con los estudiantes usando técnicas de minería clustering y de reglas de asociación	34
2.2.8. Resultados de análisis de datos de estudiantes en la Universidad ..	35

2.2.9. Descubrimiento de información oculta dentro de los datos sobre la inscripción de Estudiantes a la Universidad por medio de minería de datos .	35
2.2.10. Estudio y análisis de estudiantes desertores con minería de datos en cursos universitarios de tecnología	35
2.2.11. Predicción del éxito de los estudiantes mediante la aplicación de algoritmos de minería de datos	36
2.2.12. Aplicaciones de minería de datos en la Educación Superior	36
2.2.13. Minería de datos para conjunto de datos de alumnos - Sistema de gestión del conocimiento de maestros para la Educación Superior	36
2.2.14. Minería de datos educacional para mejora del rendimiento de estudiante	37
2.2.15. Competencia de predicción de mejoras y análisis del rendimiento de Modelo Probabilístico por medio de redes neuronal de los estudiantes	37
2.2.16. Los estilos de enseñanza/aprendizaje como soporte de la actividad docente	37
2.2.17. Software de minería de datos para el preprocesado	38
2.2.18. Software de minería de datos para la extracción de conocimientos .	40
CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN	42
3.1. MATERIALES Y METODOS	43
3.1.1. Equipos y materiales.....	43
3.1.1.1. Talento Humano	43
3.1.1.2. Recursos de software	43
3.1.1.3. Recurso de Hardware utilizados.....	44
3.1.1.4. Presupuesto	44

3.2. MÉTODOS Y TÉCNICAS UTILIZADOS EN LA INVESTIGACIÓN	45
3.2.1. Método científico	45
3.2.2. Método inductivo	45
3.2.3. Método deductivo	45
3.2.4. Método analítico	46
3.2.5. Método descriptivo	46
3.3. TIPO DE INVESTIGACIÓN.....	46
3.3.1. Diseño de investigación.....	46
3.3.2. Pasos de un Cuasi Experimento	46
3.3.3. Descripción de experimentos	47
3.4. TÉCNICAS DE INVESTIGACIÓN	48
3.4.1. Observación Directa	48
3.5. POBLACIÓN Y MUESTRA	48
CAPÍTULO IV: RESULTADOS Y DISCUSIÓN	49
4.1. RESULTADOS.....	50
4.1.1. Fuente de extracción de los datos.....	50
4.1.2. Preparación de datos (Preprocesado).....	51
4.1.2.1. Análisis de los atributos	58
4.1.2.2. Transformación y selección de atributos.....	77
4.1.2.3. Transformación de atributos	78
4.1.2.4. Vistas minables.....	79

4.1.2.5. Imputación de datos.....	83
4.1.2.6. Técnicas de discretización	84
4.1.3. Extracción de conocimiento.....	84
4.2. DISCUSIÓN	99
CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES.....	100
5.1. CONCLUSIONES	101
5.2. RECOMENDACIONES.....	102
CAPÍTULO VI: LITERATURA CITADA	103
6.1. BIBLIOGRAFÍA	104
CAPÍTULO VII: ANEXOS.....	110

ÍNDICE DE TABLAS

	PÁG
Tabla 1.- Comparación de software de preprocesado de datos	38
Tabla 2.- Comparaciones de software para la extracción de conocimiento .	40
Tabla 3.- Talento Humano	43
Tabla 4.- Recursos de Software.....	43
Tabla 5.- Recurso de Hardware	44
Tabla 6.- Presupuesto.....	44
Tabla 7.- Descripción de Atributos.....	52
Tabla 8.- Vista minable N° 1	79
Tabla 9.- Vista Minable N° 2	80
Tabla 10.- Vista minable N° 3	80
Tabla 11.- Vista minable N° 4	81
Tabla 12.- Vista minable N° 5	82
Tabla 13.-Descripción de los algoritmos de los experimentos	85
Tabla 14 Análisis del conjunto de reglas de algoritmos difusos	87
Tabla 15 Análisis de las desviaciones estándar.....	89
Tabla 16 Indices de precisión de los algoritmos difusos	90
Tabla 17.- Resultados de los algoritmos aplicados en Keel.....	92
Tabla 18.- Base de reglas	93

ÍNDICE DE ILUSTRACIÓN

	PÁG
Ilustración 1.- Ejemplo de Kmeans	27
Ilustración 2.- Software Keel	31
Ilustración 3.- Software Weka	31
Ilustración 4.- Atributo de carrera.....	59
Ilustración 5.- Atributo Sexo	59
Ilustración 6.- Atributo primer año	60
Ilustración 7.- Atributo tiempo de aprobación.....	60
Ilustración 8.-Atributo Índice de aprobación de primer año	61
Ilustración 9.- Atributo experiencia docente de primer año	61
Ilustración 10.- Atributo Índice académico de primer año	62
Ilustración 11.- Atributo promedio de segundo año.....	62
Ilustración 12 Atributo tiempo de aprobación de segundo año	63
Ilustración 13 Atributo Índice de aprobación de segundo año	63
Ilustración 14 Experiencia docente de segundo año.....	64
Ilustración 15 Atributo Índice académico de segundo año	64
Ilustración 16 Atributo: promedio de tercer año	65
Ilustración 17 Atributo: Tiempo de aprobación de tercer año.....	65
Ilustración 18 Índice de aprobación de tercer año	66
Ilustración 19 Atributo experiencia docente de tercer año	66

Ilustración 20 Atributo índice académico de tercer año	67
Ilustración 21 Atributo promedio de cuarto año	67
Ilustración 22 Atributo tiempo de aprobación de cuarto año	68
Ilustración 23 Atributo Índice de aprobación de cuarto año	68
Ilustración 24 Atributo experiencia docente de cuarto año	69
Ilustración 25 Atributo Índice académico de cuarto año.....	69
Ilustración 26 Atributo promedio de quinto año.....	70
Ilustración 27 Atributo tiempo de aprobación de quinto año	70
Ilustración 28 Atributo Índice de aprobación de quinto año	71
Ilustración 29 Experiencia docente de quinto año.....	71
Ilustración 30 Atributo Índice académico de quinto año.....	72
Ilustración 31 Atributo hábitos de estudio	72
Ilustración 32 Estrategia de estudio	73
Ilustración 33 Atributos Estilo de aprendizaje	73
Ilustración 34 Atributo auditivo	74
Ilustración 35 Atributo Lectura	74
Ilustración 36 Atributo Visual.....	75
Ilustración 37 Atributo Kinésico.....	75
Ilustración 38 Atributo situación	76
Ilustración 39.- Análisis de atributos en weka	77
Ilustración 40.- Selección de atributos en weka.....	78

Ilustración 41.- Imputación de datos FKmeans	83
Ilustración 42.- Discretizar datos en Keel.....	84
Ilustración 43.- Análisis de reglas	87
Ilustración 44.- Análisis de la desviación estándar.....	88
Ilustración 45.- Análisis de precisión.....	89
Ilustración 46.- Resultados	91

RESUMEN EJECUTIVO

La presente tesis consiste en representar conocimiento oculto a través de sistemas difusos para evaluar los estilos de enseñanza/ aprendizaje entre estudiantes y docentes de la Universidad Técnica Estatal de Quevedo.

Mediante la ingente cantidad de información almacenada en las bases de datos de la institución educativa surge la creación del tema de investigación, para la extracción del conocimiento oculto de la misma.

El objetivo principal del presente trabajo es encontrar un modelo basado en lógica difusa que permita representar los estilos de enseñanza / aprendizaje entre docentes y discentes respectivamente.

Se pretende contribuir con la Unidad de Planeamiento Académico proporcionando un conjunto de reglas que permita modelar la situación de estudio en el presente trabajo y que a su vez pueda servir como un aporte para el proceso de toma de decisiones de la unidad antes mencionada.

En el desarrollo de la tesis se efectuó varios experimentos con el conjunto de datos que se contó para así aplicar los diferentes algoritmos de lógica difusa por medio del software keel para la extracción del conocimiento oculto del problema de estudio.

Los resultados expuestos en el presente trabajo fueron seleccionados en base a los niveles de confiabilidad y cobertura de cada uno de los algoritmos aplicados en el caso de estudio permitiendo así modelar la relación entre los diferentes estilos de enseñanza / aprendizaje de los docentes y discentes de la UTEQ, quedando a disposición de la Unidad de Planeamiento Académico como apoyo en el proceso de toma de decisiones.

ABSTRACT

This thesis is to represent hidden knowledge through fuzzy systems to evaluate teaching styles / learning among students and teachers of the State Technical University of Quevedo.

By the vast amount of information stored in databases of the school comes the creation of the research topic for the extraction of hidden knowledge of it.

The main objective of this work is to find a system based on fuzzy logic that allows represent the styles of teaching / learning among teachers and students respectively model.

It aims to contribute to the Academic Planning Unit providing a set of rules that allow modeling the situation of study in this paper and which in turn can serve as a contribution to the decision-making process of the aforementioned unit.

In the development of theses various experiments with the dataset that was counted in order to apply different fuzzy algorithms keel through software to extract the knowledge of the problem of hidden study was conducted.

The results presented in this study were selected based on the levels of reliability and coverage of each of the algorithms applied in the case study allowing to model the relationship between the different styles of teaching / learning of teachers and students of the UTEQ and made available to the Academic Planning Unit as support in the decision making process.

CAPÍTULO I: MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1. INTRODUCCIÓN

Las personas se diferencian en la forma de percibir, pensar, procesar la información, sentir y comportarse. En las aulas de clases estos hechos son evidentes poniéndose de manifiesto que las diferencias de estilo afectan a muchos aspectos del proceso de enseñanza/aprendizaje y a la toma de decisiones académica y profesional.

Los estilos de enseñanza que apuntan a la tendencia de los docente por adoptar un determinado modo de relacionarse con el alumno en función de percibir las necesidades, de las demandas específicas de la tarea, intereses, capacidades del alumno y de pensar acerca de su actividades educativa (González Peiteado, 2013)

El análisis inteligente de datos está abarcando importantes avances en la sociedad de manera especial en el ámbito educativo, las TIC¹ han generado la necesidad de que actualmente sea habitual que las organizaciones dispongan de repositorio de datos con enormes cantidades variables, permitiendo almacenar información útil para la organización o institución educativa que lo requiere. En pocas ocasiones, el tamaño de las bases de datos hace parecer usual que se encuentren con una gran heterogeneidad en las mismas recogida de la diversidad de información tanto textual como numérica, fruto del interés de las organizaciones por disponer de cuanta información sea posible (Martínez & López, 2010)

En la presente investigación se utiliza para el análisis de datos y descubrir características, relaciones, tendencias o dependencias de los datos desconocidas se emplea minería de datos en la actualidad (Rob & Coronel, 2004)

La evolución experimentada en el campo de extracción de conocimiento de base de datos, fruto de la sinergias ofrecidas por las continuas investigaciones

¹ Tecnología de la información y comunicación

interdisciplinarios desarrollada en áreas como la gestión de base de datos, el aprendizaje automático, razonamiento de patrones, inteligencia artificial, ha mejorado sosteniblemente las posibilidades para obtener conocimiento de las base de datos con las que actualmente trabajan las organizaciones (Martínez & López, 2010)

Según Vialardi (2011), la realización de un estudio que permite analizar más allá del rendimiento de los estudiantes basados en las calificaciones obtenidas, sino más bien con un enfoque hacia la interpretación de las diversas formas de aprendizaje de ellos, los estilos de enseñanza que imparten los docentes y los estilos de aprendizaje con lo que los alumnos asimilan los nuevos conocimientos, puede ayudar a apreciar algunas de las razones que inciden en la deserción y el aprovechamiento de los alumnos universitarios.

Los estudiantes aprenden de formas desiguales: razonando lógicamente e intuitivamente, viendo y escuchando, construyendo analogías y modelos matemáticos, reflexionando y actuando, memorizando y visualizando, (Duran & Costaguta, 2007) de la misma forma los docentes imparten sus clases de diversas maneras.

Además analiza que los métodos de aprendizajes del estudiante de alguna manera pueden estar limitados a su habilidad y los conocimientos que tenga, expresando también que los métodos de enseñanza dependen de cada uno de los educadores. Con lo aquí argumentado se puede inferir que el aprendizaje de un alumno podría estar relacionado con la compatibilidad de sus estilos de enseñanza / aprendizaje. (Duran & Costaguta, 2007)

En la actualidad se desconoce de un estudio de investigación en las universidades ecuatorianas que analice los estilos de enseñanza / aprendizaje de la Universidad Técnica Estatal de Quevedo, la misma que sea de utilidad para el desarrollo académico y seguimiento de la forma como enseñan aprenden los docentes y estudiantes respectivamente.

La presente investigación se basará en los estilos de enseñanza / aprendizaje VARK² para la obtención de los experimentos que permitan analizar las reglas obtenidas como resultados.

La elaboración de un modelo que permita detectar reglas que puedan inferir en la tendencia de aprobación o reprobación de los estudiantes podría servir como apoyo en el proceso de toma de decisiones de la gestión académica de la institución educativa, encaminado al mejoramiento de los índices de calidad institucional en el ámbito académico.

² Visual, Auditivo, Lectura, Kinésico

1.2. JUSTIFICACIÓN

El propósito de efectuar un estudio absoluto de los datos de la Universidad Técnica Estatal de Quevedo para brindar a los gestores académicos un soporte idóneo, para mejorar las estrategias de estudios y encuadrar sus ámbitos de preparación académica, los cuales conjuntamente con la preparación y experiencia de enseñanza del docente de la Universidad además de la información de los estudiantes aporten significativamente a la realización del trabajo.

La acción de reprobación o desertar en los estudios podría no ser solo ocasionada por los bajos rendimientos sino más bien por la presencia de otros factores internos y externos, dando lugar a la realización de una investigación que permita analizar más allá de las calificaciones.

En la Universidad Técnica Estatal de Quevedo se cuenta con un enorme repositorio de datos que servirían para realizar dicha investigación creando consigo la necesidad de crear un modelo basado en lógica difusa que permita detectar un conjunto de reglas para inferir la tendencia de aprobación o reprobación de los estudiantes.

La creación de un modelo difuso que se adapte al análisis de los datos obtenidos sobre las diversas formas de enseñanza que imparten los docentes, así como de los estilos de aprendizajes que utilizan los alumnos coadyuvaría en el mejoramiento de la gestión académica de la UTEQ.

Mediante del análisis de las reglas que se obtengan se podrá observar el conocimiento oculto de las bases de datos de la institución, permitiendo así encontrar la relación de los estilos de enseñanza / aprendizaje entre estudiantes y docentes.

1.3. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA

1.3.1. Análisis del problema

Ante la complejidad de analizar el rendimiento de los estudiantes cuantitativamente y de tratar de comprender las ventajas y desventajas que existen en los estilos de enseñanza / aprendizaje utilizados en la Universidad Técnica Estatal de Quevedo podrían ocasionar deserciones, reprobaciones por parte de los estudiantes.

La gran cantidad de estudiantes con los que cuenta la UTEQ ocasiona complejidad en el análisis de los indicadores que representan las diversas formas de aprendizaje de los estudiantes, así mismo de los estilos de enseñanza del docente; recalcando que todos tienen diversas formas al momento de impartir/ receptor la clase.

En las Universidades de Ecuador y de manera especial en la UTEQ no se cuenta con una línea de investigación que se haya enfocado en el avance académico de los estudiantes por lo cual no existe un modelo que defina las variables de los estilos con los cuales enseñan los docentes y a su vez con los que los estudiantes adquieren conocimiento.

La ingente cantidad de datos con que se cuenta podrían ser utilizados en un modelo que sirva como apoyo en el análisis de los datos almacenado en las bases de datos de la institución, la cual podría ser un aporte en el proceso de toma decisiones de la gestión académica.

1.3.2. Formulación

Mediante lo abarcado en la sección anterior se plantea la siguiente interrogante: ¿Se encontrará un modelo basado en lógica difusa que permita representar los estilos de enseñanza/ aprendizaje entre los estudiantes y docentes?

1.3.3. Sistematización

- ✚ ¿Cuáles son las variables que se ajustan al proceso de estilos enseñanza/aprendizaje utilizados por los docentes y estudiantes de la UTEQ?
- ✚ ¿Se logrará aplicar diferentes técnicas que aporten a la creación de las reglas difusas que describa las relaciones entre las variables de decisión y el rendimiento estudiantil?
- ✚ ¿Cuál será el modelo que permita analizar mejor los estilos de enseñanza/aprendizaje de los docentes y estudiantes?

1.4. OBJETIVOS

1.4.1. Objetivo General

Encontrar un modelo basado en lógica difusa, que permita representar conocimiento sobre los estilos de enseñanza/aprendizaje utilizados por los docentes y estudiantes, para la ayuda de toma de decisión de la gestión académica de la Universidad Técnica Estatal de Quevedo.

1.4.2. Objetivos Específicos

-  Identificar las variables que aporten significativamente en el análisis de los datos referentes a los estilos de enseñanza/aprendizaje de los estudiantes y docentes.
-  Aplicar diferentes técnicas para encontrar reglas difusas que describan las relaciones entre las variables de decisión y el rendimiento estudiantil.
-  Seleccionar el modelo que permita analizar mejor los estilos de enseñanza/aprendizaje de los docentes y estudiantes.

1.5. HIPÓTESIS

1.5.1. Planteamiento

Las técnicas basadas sistemas difusos permitirán obtener una descripción más detallada de la influencia de los diferentes estilos de enseñanza/ aprendizaje sobre el rendimiento académico de los estudiantes de la Universidad Técnica Estatal de Quevedo.

1.5.2. Matriz de Conceptualización

TIPOS DE VARIABLE	DEFINICIÓN CONCEPTUAL	DIMENSIÓN DE LA VARIABLE	INDICADORES
VARIABLE INDEPENDIENTE Algoritmo de sistemas difusos	Se adapta mejor al mundo real en el que vivimos, e incluso puede comprender y funcionar con nuestras expresiones	Modelación	✓ Estrategia de estudio ✓ Estilos de aprendizajes ✓ Estilos de enseñanza del docente ✓ Habito de estudio ✓ Índice de aprobación

VARIABLE DEPENDIENTE Análisis de las relaciones de los estilos de enseñanza / aprendizaje entre estudiantes y docentes	Resultados obtenidos luego de haber analizado cada uno de los modelos	Eficacia	✓ Nivel de precisión ✓ Número de reglas ✓ Desviación estándar ✓ Conjunto de reglas de mejor precisión .
--	---	----------	---

CAPÍTULO II: MARCO TEÓRICO DE LA INVESTIGACIÓN

2.1. FUNDAMENTACIÓN CONCEPTUAL

2.1.1. Extracción de conocimiento

Extracción no trivial de información implícita, desconocida a prior, y potencialmente útil de los datos. (Frawley, 1992)

KDD es una disciplina reciente, aun en emersión, e interdisciplinar que combina de forma sinérgica métodos de áreas como el aprendizaje automático, el reconocimiento de patrones, estadística, base de datos y visualización para extraer conocimiento de un conjunto de datos (base de datos), con ampliaciones muy diversas, tales como medicina, biología, finanzas, marketing o astronomía. (Fayyad & Simoudis, 1997)

2.1.2. Estilo de aprendizaje

Se puede definir a los estilos de aprendizaje como los rasgos cognitivos, fisiológicos y afectivos, de cómo los alumnos perciben interacciones y responden a sus ambientes de aprendizaje (Jimenez, 2008)

2.1.3. Estrategias de Aprendizaje

García (2012) define las Estrategias de Aprendizaje como “actividades que se eligen con el propósito de facilitar la adquisición o las secuencias integradas de procedimientos o, el almacenamiento y/o la utilización de información o conocimientos”. Pozo (1996) explica lo que son las estrategias de aprendizaje y los procesos que están relacionados.

Marqués (2001) considera que la estrategia didáctica con la que el profesor pretende facilitar los aprendizajes de los estudiantes, integrada por una serie de actividades que contemplan la interacción de los alumnos con determinados contenidos. La estrategia didáctica con que se eduque a los estudiando debe

proporcionar: motivación, información y orientación para realizar sus aprendizajes, teniendo en cuenta algunos principios:

- ✓ Razonar las características de los estudiantes: estilos de aprendizaje y cognitivos.
- ✓ Razonar las motivaciones e intereses de los estudiantes.
- ✓ Organizar en el salón de clases: el espacio, los materiales didácticos, el tiempo, juegos recreativos.
- ✓ Proporcionar la información necesaria cuando sea preciso: monografías, libros, periódicos, apuntes, páginas web, asesores.
- ✓ Recurrir a metodologías activas en las que se aprenda practicando.
- ✓ Razonar un adecuado tratamiento de los errores que sea punto de partida de nuevos aprendizajes.
- ✓ Prever que los estudiantes puedan controlar sus aprendizajes.
- ✓ Razonar actividades de aprendizaje colaborativo, pero tener presente que el aprendizaje es individual.
- ✓ Realizar una evaluación final de los aprendizajes. (García, J. L.; Sánchez, C.; Jiménez, M. A.; Gutiérrez, M, 2012)

A continuación se citan varios estilos de aprendizaje:

- ✓ *Preferencias del entorno de aprendizaje*
- ✓ *Preferencias emocionales*

✓ *Preferencias sociales*

✓ *Preferencias fisiológicas*

✓ *Preferencias psicológicas* (Arnold & Blessie, 2013)

2.1.4. Estilo de enseñanza

Los Estilo de enseñanza que apunta a la tendencia de los catedráticos de adoptar un determinado modo de interactuar con el alumno en función de las tarea, de percibir las necesidades, intereses, capacidades del alumno y de razonar acerca de su praxis educativa (González Peiteado, 2013)

2.1.5. Estilo de enseñanza modelo VARK

Se los define como una forma propia de interaccionar con los alumnos y que se manifiesta tanto en la decisiones preactivas, durante las decisiones interactivas y en las decisiones postactivas (Sicilia & Delgado, 2002)

Entre los modelos que ofrecen descripciones sobre las preferencias de aprendizaje se encuentran los estilos de aprendizaje de Gregorc, , el modelo enseñanza-aprendizaje de Felder-Silverman, el de Dunn Dunn, la teoría de aprendizaje de Kolb, etc (Hawk & Shah, 2007).

En el presente trabajo utilizamos el modelo VARK, propuesto por Neil Fleming y Colleen Mills (2006), consiste en la determinación de procesar la información desde el punto de vista sensorial de los alumnos. El modelo VARK parte del supuesto de que si pueden identificar su propio estilo los estudiantes, entonces pueden adecuarse a los estilos de enseñanza de sus profesores y actuar sobre su propia modalidad en un intento por incrementar el aprovechamiento en su aprendizaje (Lozano R., 2001).

VARK es el acrónimo formado por las letras iniciales (en inglés): Visual (V), Auditiva (A), Lectoescritura (R) y Kinestésica (K). Dichas categorías se pueden describir de la siguiente forma:

Visual (visual): preferencia de representar la información por maneras gráficas y simbólicas

Lectoescritura (read/write): inclinación por la información impresa en forma de palabras.

Auditivo (aural): inclinación por escuchar la información.

Kinésico (kinesthetic): preferencia perceptual relacionada con el uso de la experiencia y la práctica, ya sea real o simulada. (Fleming & D. Baume, 2011)

En consecuencia, el modelo VARK provee una cuantificación de las preferencias de los estudiantes en cada una de las cuatro modalidades sensoriales mencionadas.

Llegando a deducir que los estudiantes visuales prefieren aprender por medio de gráficas, mapas, cuadros, diagramas, cuadros de flujo, diferentes colores, fotografías, figuras y diferentes arreglos espaciales.

Los auditivos aprenden explicando nuevas ideas a otros, en discusiones entre profesores y estudiantes, etc.

Los lectoescritores optan por ensayos, listas, manuales, reportes, libros de texto, definiciones, lecturas, páginas web y a través de tomar notas, mientras que los kinésicos aprenden mediante viajes, ensayos de prueba y error, manipulación de artefactos, recetas, empleo de laboratorios y soluciones a problemas, uso de sus sentidos y haciendo colecciones de muestras (Hawk & Shah, 2007)

2.1.6. Estilo de aprendizaje VARK

El equipo de trabajo conformado por el profesor Neil Fleming y Collen Mills de la Universidad de Lincoln, en Nueva Zelanda, desarrollan en el año 1992 una propuesta para clasificar a las personas de acuerdo a su preferencia en la modalidad sensorial a la hora de procesar información o contenidos educativos. Los autores de la investigación consideran que las personas reciben información a través de los sentidos y que el cerebro selecciona parte de esa información e ignora el resto. Las personas prestan atención en función de sus intereses, seleccionando la información, pero también influye cómo se recibe la información.

Surge un instrumento que más que una herramienta de diagnóstico, pretendía ser un catalizador para la reflexión y análisis de “cómo aprendo más rápido y mejor”, “en cuáles condiciones”, a este instrumento se le denominó VARK que es el acrónimo de las cuatro letras iniciales correspondientes a las preferencias modales sensoriales (Visual, Aural, Read/Write, Kinesthetic por sus siglas en inglés) (Visual, Auditivo, Lectura/Escritura y Quinestésico), a continuación se describe cada una: (Orientación Andujar, 2014)

Visual (visual): preferencia por maneras gráficas y simbólicas de representar la información.

Lectura – escritura (read/write): preferencia por información impresa en forma de palabras.

Auditivo (aural): preferencia por escuchar la información.

Kinésico (kinesthetic): preferencia perceptual relacionada con el uso de la experiencia y la práctica, ya sea real o simulada. (Orientación Andujar, 2014)

2.1.7. Algoritmo

Se define como algoritmo en matemáticas, lógica, ciencias de la computación y disciplinas relacionadas, un algoritmo es un conjunto reglas bien definidas o prescrito de instrucciones, ordenadas y finitas que permite realizar mediante procesos sucesivos una actividad que no formen dudas a quien deba realizar dicha actividad. (Brassard, Gilles; Bratle, Paul, 1997)

2.1.8. Sistema Difuso

Un sistema experto difuso es un sistema basado en reglas con variables lingüísticas³ que emplea un mecanismo de inferencia difusa. Un sistema experto difuso es un sistema basado en reglas con variables lingüísticas que emplea un mecanismo de inferencia difusa. (Ascolano, Cazorla, Alfonso, Colomina, & Lozano, 2003)

Un sistema basado en lógica difusa (Fuzzy Logic System, FLS) se basa en un modelo no lineal de un vector de entrada en una salida escalar. Un FLS puede expresarse como una combinación lineal de funciones, base difusas y es un aproximador universal no lineal de funciones, propiedad que comparte con el perceptrón multicapa. El desarrollo en base difusas es muy fornido porque se pueden obtener a partir de datos numéricos o bien conocimiento lingüístico; se pueden presentar en la forma de reglas IF-Then. Existen dos tipos de conocimientos: (Martín, 2000)

- Conocimiento Objetivo: cuantificada diariamente mediante modelos matemáticos.
- Conocimiento Subjetivo: posee información lingüística la cual no es posible cuantificar mediante el uso de modelos matemáticos tradicionales

³ En el marco de la lógica difusa los términos usados difusamente, es decir inexactos, vagos o inciertos, son llamados variables lingüísticas (Martínez & López, 2010)

A la hora de enfrentarnos a un problema el conocimiento subjetivo suele ser ignorado, sí es que para evaluar las soluciones que se le da se toma en cuenta. la lógica difusa o Fuzzy Logic (FL) sugiere la posibilidad de utilizar ambos tipos de conocimiento para resolver problemas reales (Martín, 2000)

Aparecen dos posibles formas de abordar el problema:

- **Basada en Modelos:** se representa por modelos matemáticos la información objetiva, y por afirmaciones lingüísticas convertidas en reglas la información subjetiva, que son cuantificadas usando FL.
- **Libre de Modelos:** de los datos numéricos y de la combinación con información lingüística (ofrecida por el experto) usando FL se obtienen las reglas. (Martín, 2000)

2.1.9. Lógica Difusa

Como la teoría del razonamiento incierto, impreciso o aproximado se puede estar definiendo a la lógica difusa. (Álvarez, 1994)

Otra definición plantea que: “Los ordenadores pueden diluir del blanco y el negro de la lógica ordinaria en los grises con que el sentido común percibe un mundo incierto gracias a que la lógica borrosa es una rama de la inteligencia artificial” (Puente, 1996)

La restricción única de la lógica borrosa es que los grados de pertenecía de un objeto a conjuntos complementarios deben sumar uno. Si el aire parece fresco en un 30%, también deberá parecer no fresco en un 70% (Puente, 1996)

La lógica borrosa consta de reglas de la forma "si..., entonces...", que transforman entradas en salidas, en un conjunto borroso en otro (Puente, 1996)

2.1.10. Variables Difusas

Las variables difusas son variables en las que sus valores se representan mediante términos lingüísticos. Las variables difusas proporcionan la capacidad para expresar, trabajar con observaciones y medidas de incertidumbre, que son más ajustadas a la realidad que las variables con cambios escabrosos (Soler, 2007).

2.1.11. Preprocesamiento de Datos

El preprocesamiento de los datos, es la fase donde se ocupa la mayor parte del tiempo para analizar y preparar los datos y por tal motivo es muy importante. La información que se va formando del proceso de limpieza y el preprocesamiento de los datos nos van generando información o datos muy valiosos que pasando por cada proceso los resultados esperados nos llevan a tomar excelentes decisiones (Dais-ujat, 2007)

La importancia de la preparación de los datos se ve reflejada en los tres aspectos siguientes:

- Los datos del mundo real puede ser incompletos, inconsistente, o presentar ruido.
- La preparación genera conjuntos de datos menores que el conjunto original, lo cual puede mejorar significativamente la eficiencia del algoritmo MDD (Minería de datos).
- La preparación da lugar a datos de calidad, al recuperar instancias incompletas, corregir errores o resolver conflictos (Dais-ujat, 2007)

Para mejorar la calidad de los conjuntos de datos, desarrollando cualquiera de las tareas anteriormente citadas.

- Limpieza de datos: se aumenta la calidad de los datos al nivel requerido mediante técnicas de análisis selectivo. Esta técnica consiste en la eliminación de datos erróneos e inconsistentes.
- Reducción de datos (RDD): Consiste en decidir qué datos deben ser utilizados para el análisis. (Dykinson, 2008)
- Transformación de datos: se basa en que es necesario que los datos sean enriquecidos con otras fuentes de información ya sean internas o externas para reducir los datos y el número de variables y transformarlos en un formato que ayude al mejoramiento de la información para llegar al conocimiento requerido.

Una de las tareas o puntos que puede abarcar del preprocesamiento es:

- Derivar nuevos atributos: que consistes en crear nuevos campos explícitos con relaciones entre atributos conocidos que puede ser el análisis más sencillo.
- Agrupamiento: Este consiste en donde hay relaciones uno a muchos y podemos convertir estas relaciones en uno a uno para perfeccionar un mejor análisis y encontrar un mejor resultado. (Dais-ujat, 2007)

2.1.12. Discretización de características

Los atributos aparecen en las instancias en muchos casos en diferentes formatos: nominales, discretos y continuos. La necesidad de discretizar los atributos continuos o con un número alto de valores discretos puede venir dispuesta por el algoritmo de aprendizaje que se emplea sobre ellos, que o bien no puede aplicarse sobre atributos continuos o bien es altamente ineficiente.

La discretización podríamos definirla como el proceso de cuantificar atributos continuos y podemos clasificar las diferentes técnicas de la siguiente forma:

- **Combinación:** se sigue una estrategia de abajo hacia arriba. Consiste en comenzar con la lista completa de valores continuos utilizándolos como puntos de corte e ir eliminándolos mediante combinaciones sucesivas entre ellos conforme el proceso de discretización progresa.
- **División:** la estrategia seguida en este conjunto de técnicas es de arriba hacia abajo. De esta forma, se comienza con un conjunto vacío de puntos de corte y se van añadiendo nuevo por división del espacio continuo mientras el proceso progresa. Los diferentes métodos podemos clasificarlos según su naturaleza:
 - ✓ **No supervisada:** las técnicas no emplean la clase a la que pertenecen las instancias para discretizarlas. Las técnicas que aquí se agrupan comparten su medida de discretización, siendo por acumulación. Consiste en discretizar los atributos continuos asignándoles un determinado número de acumuladores o contenedores.
 - ✓ **Supervisada:** se caracterizan por utilizar información sobre la clase a la que pertenece una instancia durante el proceso de discretización (Dykinson, 2008).

2.1.13. KDD: Proceso de Extracción de conocimiento

La Extracción de conocimiento está ligado con el proceso de descubrimiento conocido como *Knowledge Discovery in Databases* (KDD⁴), el cual se refiere al proceso de descubrir conocimiento e información útil dentro de los datos contenidos en algún repositorio de información (Han & Kamber, 2001).

⁴Proceso de extracción de Conocimiento

En la extracción del conocimiento se detallan los siguientes pasos:

- **Limpieza de datos** (para eliminar los datos inconsistentes y el ruido de los repositorios de datos)
- **Integración de datos** (múltiples fuentes de datos pueden ser combinadas)
- **Selección de datos** (los datos pertinentes al análisis se recuperan de la base de datos)
- **Transformación de datos** (donde los datos se consolidan en formas apropiadas para la minería o transforman)
- **Minería de datos** (Aplicación de métodos inteligentes para extraer patrones de datos)
- **Evaluación del patrón** (identifican los patrones significativos que representan el conocimiento sobre la base de algunas medidas de intereses)
- **Presentación Conocimiento** (se utilizan técnicas de representación y visualización del conocimiento, para presentar el conocimiento extraído para el usuario) (Han & Kamber, 2001)

2.1.14. Extracción de conocimiento en grandes volúmenes de datos

KDD es el proceso no trivial que trata obtener patrones válidos, nuevos, útiles y comprensibles dentro de un conjunto de datos, refiriéndose como patrones válidos a los que permiten representar conocimientos reales, y a los que cumplen con el objetivo se los considera útiles y comprensibles (Bernal L. , 2010)

2.1.15. Las fases del proceso de extracción del conocimiento

La minería de datos es un proceso que se realiza en varias fases; Análisis y definición del objetivo, selección de datos, preparación de datos, minería de datos, evaluación, interpretación e integración del modelo.

Las tareas que se pueden encontrar están basadas en una o varias de las siguientes:

- **Búsqueda de similitudes y agrupar entidades parecidas:** Se refiere a un trabajo de investigación en el que se cuenta con poco dominio de la información y se quiere empezar a tener una idea más clara al respecto. (Bernal L. , 2010)
- **Clasificar objetos:** Esta parte de una situación con mayor información la cual trata de estudiar mejor las diferencias entre los grupos, aquí se encuentran diversos modelos clasificatorios como redes bayesianas, árbol de decisión (Bernal L. , 2010)
- **Predecir:** Se trata de predecir algo a suceder, tiene similitud con la clasificación (Bernal L. , 2010)
- **Describir:** A pesar de que obtener conjuntos agrupados de objetos es un primer nivel de descripción, con el conocimiento descriptivo se trata de expresar niveles de significancia entre las variables (Bernal L. , 2010)

La selección de los datos comienza con la tecnología de data warehousing (almacenes de datos), está especialmente orientada a facilitar la localización de los datos dentro de una empresa (Bernal L. , 2010)

La preparación y selección de datos abarca el 70% del esfuerzo en los proyectos de minería de datos.

Se debe asegurar varias cosas: los datos deben tener calidad, es decir, no contener ruido, los datos deben ser necesarios, estar en forma adecuada.

La limpieza de datos esta etapa se encarga de procesar los datos para eliminar los que sean erróneos o redundantes. También se puede incluir en esta fase el paso firme de eliminar algunos de los atributos de los datos, lo que se conoce como selección de atributos y cuyo propósito es asegurar que con menos datos se pueden obtener modelos de la misma calidad y con reglas de fácil interpretación (Bernal L. , 2010)

El hecho de tener datos recogidos a mano o procedentes de la fusión de varias bases de datos suele mostrar distorsiones importantes, entre lo que se destaca:

- **Datos incompletos:** La mayor parte de estos atributos corresponden aquellos que al momentos de guardarlos no eran obligatorios (Bernal L. , 2010)
- **Datos redundantes:** A veces se repiten duplas que corresponden al mismo objeto, de manera especial cuando se fusiona distintas bases de datos (Bernal L. , 2010)
- **Datos incorrectos o inconsistentes:** Se pueden observar cuando el tipo de valores que puede recibir una instancia no se encuentra bajo ningún parámetro que lo rijan (Bernal L. , 2010)
- **Errores de transcripción:** como mayúsculas y minúsculas en valores alfanuméricos (Bernal L. , 2010)
- **Datos envejecidos:** Cundo no se actualizan los datos adecuadamente; por ejemplo, introducir la fecha de nacimiento en vez de la edad, debido a que si no se actualiza la edad, la persona siempre tendrá la misma; en

cambio, con la fecha de nacimiento, siempre se puede actualizar los datos a la fecha actual (Bernal L. , 2010)

La transformación de datos se basa en alguno de los siguientes procesos: datos numéricos a categóricos, datos categóricos a numéricos, simplificación de valores, agrupación de valores continuos, normalización de datos, añadir una etiqueta, expansión de un atributo.

Los problemas más habituales que solemos encontrar son: Derivación de datos, fusión de datos o enriquecimiento, reducción de la dimensionalidad.

- Se puede reducir el número de registros, a través de herramientas para elegir conjuntos de datos para la agrupación de modelos.
- Podemos reducir el número de atributos por ejecutar, eliminando aquellos que se considera que no aportan información para obtener el tipo de modelo objetivo.

Aparece la cuarta fase, el data mining propiamente dicho; este tipo de proceso se lo considera como un proceso de búsqueda. Se puede generar varios modelos nuevos alternativos: los modelos que ya tienen la suficiente calidad, se los puede considerar modelo final llegando a la solución, mientras si sucede lo contrario, se tiene que elegir alguno y continuar aplicando operadores de transformación hasta que se llegue a una solución o hasta que no haya más combinaciones posibles para obtener (Bernal L. , 2010)

El proceso de elaboración y elección de los modelos, puede hacerse mediante un método batch, métodos incrementales o métodos interactivos ((Bernal L. , 2010)

La evaluación e interpretación del modelo se la considera como quinta fase; el proceso de evaluación consiste en disponer de dos conjuntos de datos que partan de un conjunto inicial: se utiliza para construir el modelo un conjunto y

el otro para evaluarlo. Aunque se disponga de un tercer conjunto, el de validación (Bernal L. , 2010)

Como una sexta fase se podría tener en cuenta la integración de los resultados en el proceso. (Bernal L. , 2010)

2.1.16. Operaciones de preparación de datos

Transformación de valores se entiende a las modificaciones dentro del tipo de valores que pueden adoptar los atributos. A continuación se detallan dos técnicas de transformación: la normalización y la discretización (Bernal L. , 2010)

La **normalización** consiste en revisar que los datos estén dentro de los rangos establecidos para cada valor es importante para la construcción de modelos como el de los vecinos más cercanos ya que si no existe la normalización en el trabajo el resultado puede ser distorsionado (Bernal L. , 2010)

La **discretización** consiste en encontrar un criterio para poder dividir los valores de un atributo en dos o más conjuntos (Bernal L. , 2010)

Discretizar datos es bueno para: reducir el coste computacional pues se trabaja sobre un conjunto de valores menor, por lo cual se reduce tiempos de respuesta, reducir la velocidad en el proceso de aprendizaje. Los datos continuos ocasiona que los modelos clasificatorios que se obtienen sean mayores(Bernal L. , 2010)

En los métodos no supervisados, se utiliza la información procedente de las observaciones, se detalla lo siguiente: partición en intervalos de la misma amplitud, obtención de intervalos de discretización de igual frecuencia, método k-mean

Podemos ver un ejemplo de un conjunto de edades:

22	22	22	27	27	28	31	31	31	40	55	55
----	----	----	----	----	----	----	----	----	----	----	----

El número de valores diferentes es $n = 6$. Supongamos que queremos obtener una partición con $k = 3$ intervalos. Entonces los intervalos iniciales están formados por n/k elementos ($6/3 = 2$) y son:

Intervalo 1	Intervalo 2	Intervalo 3	Intervalo 4	Intervalo 5	Intervalo 6
22 22	22 27	27 28	31 31	31 40	55 55

Sus medianas y errores son estas:

Intervalo	Media	Distancia media (error del intervalo)
1	22	0
2	24,5	1,666
3	27,5	0,333
4	31	0
5	35,5	3
6	55	0

El error medio de esta configuración es 1,666. La nueva configuración es:

Intervalo 1	Intervalo 2	Intervalo 3	Intervalo 4
22 22 22	27 27 28	31 31 31 40	55 55

Ilustración 1.- Ejemplo de Kmean

2.1.17. Reducción de dimensionalidad

Se trata de reducir las dimensiones de los datos conservando las características de los datos iniciales pero eliminando aquellos que no son necesarios

2.1.18. Reducción del número de atributos

Trata de encontrar un subconjunto de los atributos originales permitiendo obtener modelos con la misma calidad que los que se obtendrían utilizando todos los atributos. A este proceso se le denomina problema de selección óptima de atributos (Bernal L. , 2010)

La **Prueba de significación** compara si un atributo es relevante o no a partir de sus valores (Bernal L. , 2010)

2.1.19. Reducción del número de casos

La reducción de la dimensionalidad también consiste en eliminar el número de casos por considerar (Bernal L. , 2010). Esta etapa trata de encontrar una

muestra, un subconjunto del conjunto original de casos, que muestre un comportamiento similar.

2.1.20. Tratamiento de la falta de datos

Cuando se trabaja con minería de datos es muy común encontrarse con ausencia de valores en los atributos. Dos opciones para mejorar la falta de datos son las siguientes: Sustituir el valor que falta por la media de los valores y sustituirlo por el valor más frecuente.

Las dos alternativas pueden llegar a dar resultados diferentes, de manera especial cuando el número de valores que faltan es elevado. Cabe recalcar que estos métodos sencillos llevan a introducir sesgo en los datos. Es importante tratarlos como lo que son: ausencia de información de un atributo en un caso analizado (Bernal L. , 2010)

2.1.21. Evaluación de modelos clasificatorios

La evaluación de un modelo clasificatorio consiste en predecir la clase a que pertenecen objetos no utilizados en su elaboración (Bernal L. , 2010)

Con la presencia de grandes repositorios de datos el procedimiento consiste en disponer de dos conjuntos de datos disjuntos:

- **El conjunto de entrenamiento:** El modelo clasificatorio se construye a partir de los datos que contiene este conjunto.
- **El conjunto de prueba:** Se observa la tasa de error, o precisión del modelo obtenido (Bernal L. , 2010)

Es habitual utilizar más datos para el conjunto de entrenamiento que para el de prueba (Bernal L. , 2010)

Un clasificador es mejor si se cuenta con un mayor conjunto de entrenamiento; pero cuando mayor es el conjunto de prueba más cierto es el porcentaje de error que obtenemos. No hay existe una proporción fija del número relativo de componentes de cada subconjunto, la más utilizada es de 2/3 para el conjunto de entrenamiento y 1/3 para el conjunto de prueba.

El error de una muestra se sitúa dentro de dos desviaciones típicas de la media, lo que podría representar un acierto en el 95% de los casos. (Bernal L. , 2010)

2.1.22. Software Keel

Es una herramienta de software que facilita el análisis del comportamiento de aprendizaje evolutivo en los diferentes ámbitos del aprendizaje y pre-procesamiento de tareas, por lo que estas técnicas son de fácil administración para el usuario. Los modelos corresponden a los más conocidos y utilizados en cada metodología, como una característica evolutiva y la instancia selección (Cano JR; Herrera F; Lozano M, 2003), Evolución del aprendizaje de reglas difusas y vínculo de regla Mamdani (Alcalá, R; Alcala-Fdez, J; Casillas, J; Cordon, O; Herrera, F, 2006), redes neuronales artificiales genéticos (Martínez-Estudillo A; Martínez-Estudillo F; Hervás-Martínez C; García-Pedrajas N, 2006), de aprendizaje Sistemas Clasificadores (Bernadó-Mansilla E; Ho TK, 2005)etc.

La versión actualmente disponible de keel consiste en los siguientes bloques de función:⁵

- **Manejo de Datos:** Esta parte se compone de un conjunto de herramientas que se puede utilizar para construir nuevos datos, para la exportación e importación de datos en otros formatos hacia o desde Keel formato, edición de datos y visualización, para aplicar transformaciones y particionamiento a los datos, etc (J. Alcalá-Fdez; L. Sánchez; S.

⁵ <http://www.keel.es/software/prototypes/version1.0//ManualKeel.pdf>

García; M.J. del Jesus; S. Ventura; J.M. Garrell; J. Otero; C. Romero; J. Bacardit; V.M. Rivas; J.C. Fernández; F. Herrera, 2009)

- **Diseño de Experimentos (módulo fuera de línea):** El objetivo de esta parte es el diseño de la experimentación deseada sobre el seleccionado conjunto de datos y modo de muchas opciones en diferentes áreas: tipo de validación, el tipo de aprendizaje (Clasificación, regresión, el aprendizaje no supervisado), etc (J. Alcalá-Fdez; L. Sánchez; S. García; M.J. del Jesus; S. Ventura; J.M. Garrell; J. Otero; C. Romero; J. Bacardit; V.M. Rivas; J.C. Fernández; F. Herrera, 2009)
- **Experimentos Educativos (módulo on-line):** Con una parecida estructura a la parte anterior, esto permite el diseño del experimento que se puede ejecutar paso a paso con el fin de mostrar el proceso de aprendizaje de un determinado modelo mediante el uso de la herramienta de software con fines educativos.

Con todos estos bloques de función, podemos afirmar que Keel puede ser útil por diferentes tipos de usuarios, cada uno de los cuales puede esperar encontrar características en un software de DM. (J. Alcalá-Fdez; L. Sánchez; S. García; M.J. del Jesus; S. Ventura; J.M. Garrell; J. Otero; C. Romero; J. Bacardit; V.M. Rivas; J.C. Fernández; F. Herrera, 2009)



Ilustración 2.- Software Keel

2.1.23. Weka

Weka es una colección de algoritmos para tareas de minería de datos, a través de herramientas para los datos: pre-procesamiento, clasificación, regresión, clustering, reglas de asociación, y la visualización. Los algoritmos se pueden aplicar a un conjunto de datos, también es muy adecuado para el desarrollo de nuevos sistemas de aprendizaje automático (University of Waikato, 2014)

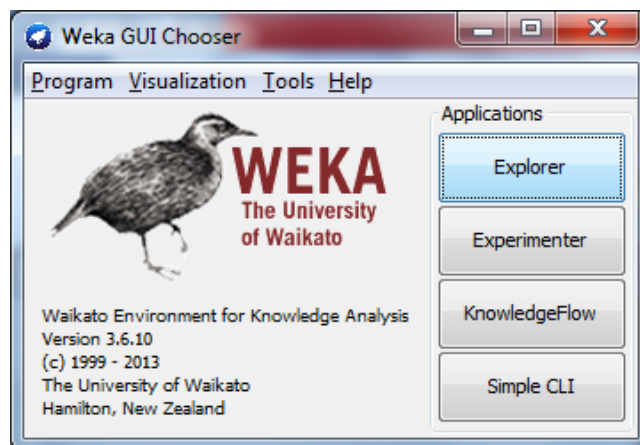


Ilustración 3.- Software Weka

2.2. MARCO REFERENCIAL

En el presente apartado constan trabajos de investigaciones de minería de datos educacional, los cuales están relacionados con la propuesta de trabajo de la tesis.

Cada uno de los trabajos citados contribuyen a la elaboración de las etapas del proceso de extracción de conocimiento aplicados en el estudio del tema central.

En el trabajo 2.2.1. , 2.2.7. y 2.2.13 se puede observar como realizan el uso del software Weka para el preprocesado de datos, debido a que es un software con una interfaz sencilla para el cliente, además de poseer una extensa colección de técnicas para preprocesamiento de datos, frente a Kmeans como se ve reflejado en la tabla 1.

Además se presenta otros trabajos con contenido de los estilos de enseñanza y aprendizaje utilizando minería de datos en otros entornos de trabajo pero orientados hacia el objetivo estudiado.

Finalmente se presenta una tabla con las ventajas y desventajas del software de minería de datos, analizando para el proceso de extracción de conocimiento.

2.2.1. Efecto de minería de datos educacional para mejorar el de aprendizaje mediante el sistema de aprendizaje electrónico

La presente investigación hace referencia al moodle, definiendo acorde al comportamiento del estudiante en el sitio 4 estilos de aprendizajes, a los resultados obtenidos se aplican algoritmos de minería de datos utilizando la herramienta WEKA ejecutando J48 y K-MEANS los cuales son algoritmos de clasificación, definiendo con K-MEANS los estilos de aprendizajes y realizan la

clasificación de los estudiantes con el algoritmo J4.8. (Chellatamilan, Ravichandran, Suresh, & Kulanthaivel, 2011)

2.2.2. Minería de datos para guiar a los estudiantes a través del proceso de inscripción basada en el rendimiento académico

Se analiza la información de los estudiantes de una universidad para las inscripciones a los cursos, se realizan 4 fases para el sistema recomendador; la primera radica en el aprendizaje automático para seleccionar los algoritmos que clasifiquen mejor los datos, la segunda para comprobar el impacto de los datos no considerados en la investigación, la tercera para calcular la menor tasa de error en la poda de los árboles de decisión, la cuarta fase para la selección del mejor algoritmo de clasificación, ya para el desarrollo de la interfaz del sistema. (Vialardi, y otros, 2011)

2.2.3. Análisis de calificación de estudiantes basados en conjuntos aproximados

Se aplica minería de datos utilizando “rough set” en los que se analiza las incidencias que tienen los cursos a los que se inscribió el estudiante, se elabora una categorización de los datos, así como también se aplica el algoritmo, para establecer las debidas ponderaciones de los cursos. (Zhi-min, Qing, & Bin, 2010)

2.2.4. Aplicación de minería por medio de reglas de asociación en el sistema de evaluación docente

En la investigación se aplicó como técnica de minería de datos las reglas de asociación de un conjunto de datos del cual se requiere saber el nivel de incidencia en las evaluaciones docentes y poder enseñar a través del desempeño. (Zhang & Gao, 2008)

2.2.5. Personalización de los procesos de aprendizaje de la minería de datos

En el presente estudio aquí espuesto se detalla la técnica "storyboarding". Mediante la aplicación de un cuestionario se obtiene los estilos de aprendizaje de los estudiantes, el cual permite determinar los estilos de acuerdo a la elección de cada uno, posteriormente aplicando la minería de datos para conseguir los patrones que representan las combinaciones de estos estilos. (Knauf, Sakurai, Takada, & Tsuruta, 2010)

2.2.6. Aplicación de minería de datos educativa para predecir las tendencias académicas y los patrones en las base de datos

En el trabajo se utilizan dos tipos de algoritmos: Apriori y el K-means, para modelar y agrupar a los estudiantes en relación a sus perfiles. El presente estudio tiene como objetivo descubrir información relevante para crear perfiles de los estudiantes en función de sus expedientes académicos, el cual puede ser apoyo para el proceso de toma de decisiones en los sistemas educativos. (Parack, Suhem; Zahid, Zain; Merchant, Fatima, 2012)

2.2.7. Extracción y análisis del desempeño docente de la gestión de retroalimentación con los estudiantes usando técnicas de minería clustering y de reglas de asociación

El artículo trata sobre el análisis del desempeño docente según la opinión de los estudiantes, donde aplican técnicas de clustering (agrupamiento), como primer punto se realiza el procesamiento de los datos con el preprocesador de datos la herramienta WEKA y posteriormente se someten estos datos a la minería de reglas de asociación. (Singh, Chandrani; Gopal, Arpita; Mishra, Santosh, 2011)

2.2.8. Resultados de análisis de datos para estudiantes en la Universidad

En el estudio se aplica un método analítico basado en el enfoque estadístico y la minería de datos (*Reglas de asociación*), con el cual se pretende mejorar la calidad de la educación universitaria, adecuando los planes de estudios en base a las puntuaciones obtenidas por estudiantes en los cursos de nivelación y saber de una forma u otra las dependencias y relaciones existentes. Los patrones o reglas descubiertas se pueden utilizar para promover efectivamente la gestión de la educación universitaria. (Xiaolong Zhang; Guirong Liu, 2008)

2.2.9. Descubrimiento de información oculta dentro de los datos sobre la inscripción de Estudiantes a la Universidad por medio de minería de datos

En el trabajo se aplican dos enfoques principales de estudio, el *descriptivo* y el *predictivo*; en la primera fase se pre-procesan los datos para reducir el número de los mismos, posteriormente utilizando el primer enfoque se realiza un análisis de tabulación cruzada (tablas de contingencia), para adicionalmente llevar a cabo un análisis de agrupamiento (clustering) para determinar la similitud entre los atributos del conjunto de datos. En cuanto al enfoque predictivo se utilizó varios modelos de predicción (como árboles de decisión, regresión, Neural Network (*Redes Neuronales*)). Los resultados pueden utilizarse como una guía o mapa de ruta para identificar qué partes de los procesos aplicados en una universidad puede ser mejorada a través de la tecnología de minería de datos y cómo la tecnología podría mejorar los procesos convencionales y de esta manera seguir siendo competitivos. (Siraj, Fadzilah; Abdoulha, Mansour Ali, 2009)

2.2.10. Estudio y análisis de estudiantes desertores con minería de datos en cursos universitarios de tecnología

En la investigación se introduce conocimiento sobre minería de datos y se selecciona el algoritmo de minería de reglas de asociación, para determinar

cuáles son las causas asociadas al buen rendimiento del estudiante universitario, y a su vez proporcionar información de soporte de decisiones para los educadores. (Zhang Z. , 2010)

2.2.11. Predicción del éxito de los estudiantes mediante la aplicación de algoritmos de minería de datos

En el trabajo se muestra resultados de una encuesta en una muestra de 907 estudiantes que en 2011/2012 año escolar han asistido la escuela secundaria en el cantón de Tuzla. La cual ha sido analizada mediante tres modelos (de árboles de decisión, Naive Bayes y Perceptrón multicapa). También se ha tomado en cuenta la estimación de las variables que han influido en el éxito de los estudiantes y los se las ha comprobado a través de cuatro filtros diferentes. Validación cruzada es el método que se utilizó en la estimación de las clasificaciones. (Osmanbegovic, Agic, & Suljic, 2014)

2.2.12. Aplicaciones de minería de datos en la Educación Superior

En el trabajo se propone el uso de técnicas de minería de datos para mejorar la eficiencia de las instituciones de educación superior. Se aplicó técnicas de minería de datos como clustering, árboles de decisión y de asociación a los procesos de enseñanza superior, la cuales ayudaron a mejorar el rendimiento de los estudiantes, la gestión del ciclo de vida, selección de cursos, para medir la tasa de retención y la administración de fondos de subvención de una institución. (Goyal & Vohra, 2012)

2.2.13. Minería de datos para conjunto pequeño de datos de alumnos - Sistema de gestión del conocimiento de maestros para la Educación Superior

En la investigación el estudio se centra en la minería de datos para conjuntos pequeños de datos de los estudiantes, para éste estudio se hace uso de dos herramientas de minería de datos (Microsoft Excel y Weka) las cuales se

utilizan para predecir la tasa de éxito que tendrán los estudiantes. (Zwilling & Natek, 2013)

2.2.14. Minería de datos educacional para mejora del rendimiento de estudiante

En el trabajo se pretende identificar a los estudiantes que necesitan asesoramiento especial por parte del docente para brindar educación de alta calidad se aplica minería de datos utilizando como métodos *Educational Data Mining (EDM)*, *ID3 Algorithm*, *Knowledge Discovery in Database (KDD)* (Priya & Kumar, 2013)

2.2.15. Competencia de predicción de mejoras y análisis del rendimiento de Modelo Probabilístico por medio de redes neuronal de los estudiantes

En el trabajo se enfoca en clasificar y evaluar la competencia comunicativa de un grupo de alumnos de un colegio de la India, se aplica como técnicas de minería de datos: Estadísticas de regresión y Redes Neuronales Probabilísticas (PNN) las cuales sirven para clasificar la categoría de rendimiento de la competencia comunicativa de los alumnos. (Ayyanathan, Kannammal, & Rekha, 2012)

2.2.16. Los estilos de enseñanza y aprendizaje como soporte de la actividad docente

En el trabajo se analiza algunos aspectos pedagógicos que son clave para optimizar el aprendizaje. Una acción práctica eficiente comienza por hacer realidad la disminución de desencuentros entre estilos de enseñanza y estilos de aprendizaje. Por tanto, es necesario que el docente conozca los modos de aproximarse al aprendizaje del alumnado, sus actitudes, valores, diferencias culturales, destrezas y hábitos de estudio. Ello conduce a asumir una praxis

pedagógica que priorice la reflexión para conseguir un cambio didáctico y un proceso de enseñanza más individualizado como medio para evitar el fracaso escolar. El fin último es elevar la calidad de la enseñanza resaltando la importancia de vincular los estilos de aprendizaje y estilos de enseñanza.

En consecuencia, se presentan los resultados de una investigación desarrollada en la Universidad de Vigo con el propósito de descubrir los estilos de enseñanza de los estudiantes de Magisterio (González-Peiteado, 2013)

2.2.17. Software de minería de datos para el preprocesado

En la siguiente tabla se observan los software para el preprocesado en la misma se muestran las ventajas y desventajas de los programas de minería de datos, permitiendo así poder llegar a la conclusión del software a aplicar en el presente trabajo de investigación.

Tabla 1.- Comparación de software de preprocesado de datos

WEKA	KNIME
VENTAJAS • Weka proporciona interfaces para la comunicación con el usuario, tal es el caso de CLI (Simple Client), esta interfaz proporciona una consola para poder introducir mandatos, posee una apariencia muy simple pero nos permite realizar tareas complejas ya que permite realizar cualquier operación soportada por Weka de forma directa; no obstante, es	VENTAJAS • Desarrollado sobre la plataforma Eclipse y programado, esencialmente, en java. Está concebido como una herramienta gráfica y dispone de una serie de nodos (que encapsulan distintos tipos de algoritmos) y flechas (que representan flujos de datos) que se despliegan y combinan de manera gráfica e interactiva.

muy complicada de manejar ya que es necesario un conocimiento completo de la aplicación. • Nos permite ubicar patrones de comportamiento de la información a procesar de tal manera que es de gran ayuda en la toma de decisiones. • Está disponible libremente bajo la licencia pública general de GNU. • Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma. • Contiene una extensa colección de técnicas para pre-procesamiento de datos y modelado. DESVENTAJAS • Existe poca documentación sobre el uso de Weka dirigida al usuario. • Un área importante que actualmente no cubren los algoritmos incluidos en Weka es el modelado de secuencia.	• Los nodos implementan distintos tipos de acciones que pueden ejecutarse sobre una tabla de datos: • Manipulación de filas, columnas, como muestreos, transformaciones, agrupaciones. • Visualización (histogramas). • Creación de modelos estadísticos y de minería de datos, como árboles de decisión, regresiones. • Scoring o aplicación de dichos modelos sobre conjuntos nuevos de datos. • Creación de informes a medida gracias a su integración con BIRT. • El carácter abierto de la herramienta hace posible su extensión mediante la creación de nuevos • Nodos que implementen algoritmos a la medida del usuario. DESVENTAJAS
---	---

	• Existe poca documentación sobre el uso de KNIME dirigida al usuario. • Interfaz de compleja interacción con el usuario.
--	--

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

2.2.18. Software de minería de datos para la extracción de conocimientos

A continuación se detallan dos software para la ejecución de algoritmos difusos, por medio de la tabla 2 se podrá llegar a tomar la decisión sobre que software utilizar en el proceso de extracción de conocimiento de los repositorios de datos de la UTEQ.

Tabla 2.- Comparaciones entre software para la extracción de conocimiento

KEEL	KNIME
VENTAJAS • Es un software para evaluar la evolución de los algoritmos de minería de datos y problemas de regresión, entre ellos: clasificación, agrupamiento y patrón de la minería. • Contiene una gran colección de algoritmos clásicos de extracción de conocimientos, técnicas de pre procesamiento (selección de instancias, selección de	VENTAJAS • Es una herramienta gráfica y dispone de una serie de nodos (que encapsulan distintos tipos de algoritmos) y flechas (que representan flujos de datos) que se despliegan y combinan de manera gráfica e interactiva. • Los nodos implementan distintos tipos de acciones que pueden ejecutarse sobre una tabla de

características, discretización, métodos de imputación de valores), Inteligencia Computacional de aprendizaje basado en algoritmos, incluido el estado evolutivo de algoritmos de aprendizaje basados en diferentes enfoques (Pittsburgh, Michigan y IRL) • Modelos híbridos como sistemas difusos genéticos, redes neuronales evolutivas, etc. • Permite realizar un análisis completo de cualquier modelo de aprendizaje en comparación con los existentes, incluido un módulo de prueba estadística para la comparación entre ellos DESVENTAJAS Existe poca documentación sobre el uso de Keel dirigida al usuario	datos: • Manipulación de filas, columnas, como muestreos, transformaciones, agrupaciones. • Visualización (histogramas). • Creación de modelos estadísticos y de minería de datos, como árboles de decisión, regresiones. • Nodos que implementen algoritmos a la medida del usuario. DESVENTAJAS • Existe poca documentación sobre el uso de KNIME dirigida al usuario. • Interfaz de compleja interacción con el usuario.
--	--

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

3.1. MATERIALES Y METODOS

3.1.1. Equipos y materiales

En la realización de este trabajo se utilizaron los siguientes recursos:

3.1.1.1. Talento Humano

Tabla 3.- Talento Humano

PERSONAL	DESCRIPCION
AUTORA	Johanna Olvera Toala
DIRECTOR DE TESIS	PhD. Amilkar Puris
ASESOR DE TESIS	Ing. Jorge Guanín

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

3.1.1.2. Recursos de software

Tabla 4.- Recursos de Software

CANTIDAD	DESCRIPCION	VALOR
1	➤ SQL Server 2008	\$ 0.00
1	➤ Keel	\$ 0.00
1	➤ LibreOffice	\$ 0.00
1	➤ Weka	\$ 0.00
1	➤ Open proj	\$ 0.00

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

3.1.1.3. Recurso de Hardware utilizados

Tabla 5.- Recurso de Hardware

CANTIDAD	NOMBRE	DESCRIPCION
1	Computador	Toshiba Características: ➤ Intel Core i3 ➤ 3 Gb.RAM ➤ 500 Gb Disco Duro ➤ CD Writer
1	Impresora	Epson L210
1	Celular	Samsung S3

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

3.1.1.4. Presupuesto

Tabla 6.- Presupuesto

CANTIDAD	MATERIAL	VALOR
	➤ Libros	\$ 00.00
	➤ Útiles de oficina	\$ 20.00
8 meses	➤ Servicio de internet	\$ 240.00
	➤ Movilización	\$ 400.00
200 min	➤ Llamadas	\$ 34.00
10 unid.	➤ Anillados	\$ 10.00
6 unid	➤ Empastados	\$ 90.00
TOTAL		\$794.00

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

3.2. MÉTODOS Y TÉCNICAS UTILIZADOS EN LA INVESTIGACIÓN

En la presente investigación los métodos a utilizar son:

3.2.1. Método científico

En el método científico se procede a la formulación de hipótesis, que en el caso implica el análisis de los estilos de enseñanza / aprendizaje entre los docentes y estudiantes de la Universidad Técnica Estatal de Quevedo basados en los estilos VARK; recolectado por medio de encuestas realizadas en la institución educativa, las cuales se encuentran almacenadas en un repositorio de datos.

El contexto supone la aplicación de la propuesta de un modelo difuso para resolver problemas planteados por la Unidad de Planeamiento Académico en cuanto estilos enseñanza/aprendizajes; además del contraste de las hipótesis con las observaciones, es decir, evaluación de la calidad de los métodos de aprendizaje por medio de las precisión, numero de reglas y la desviación estándar.

3.2.2. Método inductivo

Por medio del análisis de resultados que se obtuvieron en la realización de la tesis se puede observar como a partir del conjunto de reglas que se obtuvo se puede llegar a construir un conocimiento universal de los estilos de enseñanza/aprendizaje.

3.2.3. Método deductivo

Mediante la aplicación del método se pudo deducir los problemas que existen en los estilos de enseñanza / aprendizaje entre docentes y alumnos los cuales están relacionados a la investigación planteada, además se pudo identificar algunos factores claves que ayudarán a desarrollar el presente trabajo.

3.2.4. Método analítico

Al momento de utilizar el presente método permitirá analizar detalladamente las variables que se encuentran vinculadas directamente en el proceso de enseñar y aprender entre docentes y alumnos, así como la interpretación de las causas y efectos que genera la problemática planteada.

3.2.5. Método descriptivo

Por medio del método se trató describir el estado actual de los diferentes estilos de enseñanza/ aprendizaje en la UTEQ, no solo por sus atributos, sino más bien dando una idea completa del contexto, interpretando en forma real la investigación.

3.3. TIPO DE INVESTIGACIÓN

3.3.1. Diseño de investigación

La investigación utilizada en el presente trabajo responde a la metodología cuasi experimental, ya que se desarrolla experimentos con instancias de diferentes dimensiones para formular el conjunto de reglas difusas que describan las relaciones entre las variables de decisión y el rendimiento estudiantil. También se hace un análisis de los algoritmos que contribuyan con un alto índice de precisión y con un conjunto de reglas claras que permitan describir el objetivo del trabajo. Por último se presenta el conjunto de reglas que contribuyan a describir el contexto de la investigación.

3.3.2. Pasos de un Cuasi Experimento

Dentro de los pasos de un cuasi experimento se define el conjunto de atributos que se utilizarán para el desarrollo de la investigación, se procederá a realizar el preprocesado de datos utilizando técnicas de discretización para disminuir el

número de resultados obtenidos y de imputación de datos para completar los valores perdidos.

Se procede aplicar los algoritmos difusos para poder así medir las variables dependientes que involucran los índices de precisión, de desviación estándar y número de reglas, realizando comparaciones de las diferentes medidas de calidad que definen el modelo resultante entre los diversos algoritmos.

3.3.3. Descripción de experimentos

En la creación de experimentos se procederá a la elaboración de grupos de experimentos en cada grupo se evaluará una vista minable (conjunto de datos de la selección de atributos) a la que se aplicará cada uno de los algoritmos difusos, el cual dará como resultado un conjunto de reglas difusas que serán de-fusicadas para una mejor comprensión, así se realizará el trabajo para cada una de las vistas minables establecidas en el trabajo.

$$G_n \longrightarrow X_n, O_n$$

Dónde:

G= Grupo de vistas minables

X= Algoritmos difusos aplicados

O= Reglas difusas obtenidas

n= Número de pruebas realizadas

GRUPOS VISTAS MINABLES	ALGORITMOS	NÚMERO DE REGLAS
G ₁ Vista Minable # 1	X ₁ ALGORITMOS DIFUSOS (9)	Número de R1
G ₂ Vista Minable # 2	X ₂ ALGORITMOS DIFUSOS (9)	Número de R2

G ₃ Vista Minable # 3	X ₃ ALGORITMOS DIFUSOS (9)	Número de R3
G ₄ Vista Minable # 4	X ₄ ALGORITMOS DIFUSOS (9)	Número de R4
G ₅ Vista Minable # 5	X ₅ ALGORITMOS DIFUSOS (9)	Número de R5

3.4. TÉCNICAS DE INVESTIGACIÓN

3.4.1. Observación Directa

Por medio de la técnica se pudo analizar los índices de precisión de cada uno de los resultados que entregan los diversos algoritmos difusos, y que a su vez reduzcan la complejidad en la interpretación de reglas difusas que describan el fenómeno estudiado en la investigación.

3.5. POBLACIÓN Y MUESTRA

En el estudio del presente proyecto de investigación se trabajó en su totalidad con la población, la cual está conformada por los estudiantes de los años entre el 2002 y 2008 de la Universidad Técnica Estatal de Quevedo, además del conjunto de docentes que les impartieron clases durante sus períodos de estudios.

Debido a la disponibilidad de información por parte del SIU (Sistema de Información Universitario) se trabajó con los estudiantes del periodo antes mencionado, además cabe recalcar que el inicio de la investigación propuesta fue a partir del 2014 por lo que se tuvo que analizar solo hasta los estudiantes que ingresaron en el 2008 ya que de ellos existían los datos completos para una mejor aportación en el desarrollo del proceso de extracción de conocimiento.

La información con la que se trabaja en la investigación se encuentra en los grandes repositorios de datos de la Unidad de planeamiento académico de la universidad; se trabajará con la población completa más no con un muestreo.

CAPÍTULO IV: RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS

En el presente apartado del trabajo de investigación se podrá observar el proceso que se siguió para el descubrimiento de reglas difusas que permitan la representación de conocimiento entre los estilos de enseñanza/ aprendizaje de los docentes y discentes de la Universidad Técnica Estatal de Quevedo, por medio del uso de los siguientes software de minería de datos tales como WEKA, KEEL.

Por medio del análisis aplicado en el marco referencial en el que se puede observar que el software WEKA es uno de los que posee mayores técnicas para el preprocesado de datos y una interfaz sencilla con el usuario, se llega a la decisión de trabajar con el software mencionado en la etapa de selección de atributos.

El software de minería de datos Keel es el que se utiliza en la elaboración de la investigación para la aplicación de algoritmos difusos debido a que posee una interfaz de usuario de fácil manipulación y a que posee varios nodos difusos y permite una mejor interpretación de reglas bajo el modelo If- then.

4.1.1. Fuente de extracción de los datos

El desarrollo de la tesis de investigación comenzó con base a los ingentes repositorios de datos con lo que cuenta la institución educativa y que no han sido aplicados para encontrar conocimientos ocultos, es de gran importancia encontrar información referente a los diferentes estilos de enseñanza que aplican los docentes de la institución y a su vez los estilos de aprendizaje que usan los educandos de la misma, características de enseñanza/ aprendizaje tales como: lectura, auditivo, visual y kinésico, además de otros factores como si poseen o no hábitos de estudio.

El conjunto de datos utilizado en el presente trabajo se obtuvo del Sistema gestor de base de datos (MSQL2008) en la Universidad Técnica Estatal de

Quevedo, por medio de procedimientos almacenados se logra extraer los datos que servirán para el desarrollo del trabajo de investigación, los mismos que fueron obtenidos en base a encuestas realizadas a los docentes y estudiantes en cada año lectivo de manera específica entre los años 2002-2008.

Siendo el SIU (Sistema informático universitario) quien administraba la información de los periodos antes descrito convirtiéndose en la fuente primaria para la elaboración del trabajo.

4.1.2. Preparación de datos (Preprocesado)

La buena creación de un modelo depende de la calidad de datos minados con los que se cuenta además de la aplicación del algoritmo de minería, por lo cual después de la obtención de los datos es importante aplicar el preprocesado de los mismos para así poder obtener las distintas vistas minables.

La importancia del preprocesado es vital para el buen funcionamiento y término del proceso, por lo que suele ser una etapa que requiere una inversión considerable de tiempo (Martínez & López, 2010), La etapa se puede dividir en tres:

Selección.- Se debe seleccionar las variables (datos) de las bases de datos disponibles en función del problema y necesidades de la investigación definidas en la primera etapa. Normalmente, se disponen de diversas bases de datos integradas en lo que se conoce como almacenes de base de datos. (Martínez & López, 2010)

Procesado.- Su principal objetivo es asegurar la calidad de los datos que se van a utilizar. Es conveniente realizar un análisis descriptivo de las variables, distribución de frecuencias, distribución de valores, etc. Con el objeto de detectar los valores defectuosos, no pertenecientes al rango permitido y los datos ausentes (Martínez & López, 2010)

Transformación.- Los datos deben transformarse a un formato que sea procesable por el método de minería de datos que se vaya a utilizar para extraer el conocimiento (Martínez & López, 2010)

En la presente etapa de extracción de conocimiento se trabaja con el dataset obtenido por medio de los sistemas administradores de base de datos proporcionados por el SIU, en el cual se aplicó el preprocesado en base a algunas técnicas descritas anteriormente, a continuación se detallan la descripción de las variables utilizadas en el desarrollo del trabajo de investigación.

Tabla 7.- Descripción de Atributos

ITEM	VARIABLE	DESCRIPCIÓN	TIPO	VALOR
1	Carrera	Nombre de la carrera de estudio	categórica	
2	Sexo	Género del estudiante	categórica	{M,F}
3	promedio_1	Media de las notas del primer curso	numérica	{1..10}
4	taprobacion_1	Tiempo (cursos académicos) que tarda en aprobar el 1er curso, para obtener la variable taprobacion_1 se realizó una sumatoria del número de matrículas que realizó un estudiante para culminar un determinado año de estudio.	numérica	{1..3}
5	IndAprob_1	Índice de número de materias aprobadas sin suspensión en primer año (materias aprobadas/total materias del semestre), en la que el alumno que obtenga el 1 es porque aprobó sin supletorio, por cada uno de los mismos se les suma un 0.25	numérica	{1..5}
6	PEDAD1	Rango de edad de docentes menores a 35 años que impartieron clases en primer año	numérica	{1}
7	PEDAD2	Rango de edad de docentes entre 35 y 50 años que impartieron clases en primer año	numérica	{2}
8	PEDAD3	Rango de edad de docentes mayores a 50 años que impartieron clases en	numérica	{3}

		primer año		
9	PEXPEDOCEN	Índice de la experiencia docente de acuerdo al tiempo de enseñanza que impartieron clases en primer año	numérica	{1-5}
10	PlndAcademico	Índice de aporte de la Universidad relacionada con la preparación académica del docente , con el fin de que el estudiante tenga buena formación en su carrera universitaria y su rendimiento académico de primer año, el presente atributo se calcula en base a los títulos que tenga los docentes agregando un 0.25 por cada título académico		
11	promedio_2	Media de las notas del segundo curso	numérica	{1..10}
12	taprobacion_2	Tiempo (cursos académicos) que tarda en aprobar el 2do curso, para obtener la variable taprobacion_2 se realizó una sumatoria del número de matrículas que realizó un estudiante para culminar un determinado año de estudio.	numérica	{1..3}
13	IndAprob_2	Índice de número de materias aprobadas sin suspensión de segundo año (materias aprobadas/total materias del semestre), en la que el alumno que obtenga el 1 es porque aprobó sin supletorio, por cada uno de los mismos se les suma un 0.25	numérica	{1..5}
14	SEDAD1	Rango de edad de docentes menores a 35 años que impartieron clases en segundo año	numérica	{1}
15	SEDAD2	Rango de edad de docentes entre 35 y 50 años que impartieron clases en segundo año	numérica	{2}
16	SEDAD3	Rango de edad de docentes mayores a 50 años que impartieron clases en segundo año	numérica	{3}
17	SEXPEDOCEN	Índice de la experiencia docente de acuerdo al tiempo de enseñanza que impartieron clases en segundo año	numérica	{1-5}
18	SIndAcademico	Índice de aporte de la Universidad relacionada con la preparación académica del docente, con el fin de		

		que el estudiante tenga buena formación en su carrera universitaria y su rendimiento académico de segundo año, el presente atributo se calcula en base a los títulos que tenga los docentes agregando un 0.25 por cada título académico.		
19	promedio_3	Media de las notas del tercer curso	numérica	{1..10}
20	taprobacion_3	Tiempo (cursos académicos) que tarda en aprobar el 3er curso, para obtener la variable taprobacion_3 se realizó una sumatoria del número de matrículas que realizó un estudiante para culminar un determinado año de estudio.	numérica	{1..3}
21	IndAprob_3	Índice de número de materias aprobadas sin suspensión de tercer año (materias aprobadas/total materias del semestre) , en la que el alumno que obtenga el 1 es porque aprobó sin supletorio, por cada uno de los mismos se les suma un 0.25	numérica	{1..5}
22	TEDAD1	Rango de edad de docentes menores a 35 años que impartieron clases en tercer año	numérica	{1}
23	TEDAD2	Rango de edad de docentes entre 35 y 50 años que impartieron clases en tercer año	numérica	{2}
24	TEDAD3	Rango de edad de docentes mayores a 50 años que impartieron clases en tercer año	numérica	{3}
25	TEXPEDOCEN	Índice de la experiencia docente de acuerdo al tiempo de enseñanza que impartieron clases en tercer año	numérica	{1-5}
26	TIndAcademico	Índice de aporte de la Universidad relacionada con la preparación académica del docente, con el fin de que el estudiante tenga buena formación en su carrera universitaria y su rendimiento académico de tercer año, el presente atributo se calcula en base a los títulos que tenga los docentes agregando un 0.25 por cada título académico.		

27	promedio_4	Media de las notas del cuarto curso	numérica	{1..10}
28	taprobacion_4	Tiempo (cursos académicos) que tarda en aprobar el 4to curso, para obtener la variable taprobacion_4 se realizó una sumatoria del número de matrículas que realizó un estudiante para culminar un determinado año de estudio.	numérica	{1..3}
29	IndAprob_4	Índice de número de materias aprobadas sin suspensión en cuarto curso (materias aprobadas/total materias del semestre) , en la que el alumno que obtenga el 1 es porque aprobó sin supletorio, por cada uno de los mismos se les suma un 0.25	numérica	{1..5}
30	CEDAD1	Rango de edad de docentes menores a 35 años que impartieron clases en cuarto curso	numérica	{1}
31	CEDAD2	Rango de edad de docentes entre 35 y 50 años que impartieron clases en cuarto curso	numérica	{2}
32	CEDAD3	Rango de edad de docentes mayores a 50 años que impartieron clases en cuarto curso	numérica	{3}
33	CEXPEDOCEN	Índice de la experiencia docente de acuerdo al tiempo de enseñanza que impartieron clases en cuarto curso	numérica	{1-5}
34	CIndAcademico	Índice de aporte de la Universidad relacionada con la preparación académica del docente, con el fin de que el estudiante tenga buena formación en su carrera universitaria y su rendimiento académico en cuarto curso, el presente atributo se calcula en base a los títulos que tenga los docentes agregando un 0.25 por cada título académico.		
35	promedio_5	Media de las notas del quinto curso	numérica	{1..10}
36	taprobacion_5	Tiempo (cursos académicos) que tarda en aprobar el 5to curso, para obtener la variable taprobacion_5 se realizó una sumatoria del número de matrículas que realizó un estudiante para culminar un determinado año de estudio.	numérica	{1..3}

37	IndAprob_5	Índice de número de materias aprobadas sin suspensión en quinto curso(materias aprobadas/total materias del semestre) , en la que el alumno que obtenga el 1 es porque aprobó sin supletorio, por cada uno de los mismos se les suma un 0.25	numérica	{1..5}
38	QEDAD1	Rango de edad de docentes menores a 35 años que impartieron clases en quinto curso	numérica	{1}
39	QEDAD2	Rango de edad de docentes entre 35 y 50 años que impartieron clases en quinto curso	numérica	{2}
40	QEDAD3	Rango de edad de docentes mayores a 50 años que impartieron clases en quinto curso	numérica	{3}
41	QEXPEDOCEN	Índice de la experiencia docente de acuerdo al tiempo de enseñanza que impartieron clases en quinto curso	numérica	{1-5}
42	QIndAcademico	Índice relacionado con la preparación académica del docente, con el fin de que el estudiante tenga buena formación en su carrera universitaria y su rendimiento académico en quinto curso, el presente atributo se calcula en base a los títulos que tenga los docentes agregando un 0.25 por cada título académico.		
43	habito_estudios	¿Posee hábitos de estudios?	categorica	{1.SI, 2.NO}
44	estrategia_estudio	¿Qué Estrategia utiliza para estudiar?	categorica	1. NC=. No contesta 2. ICT= Interpretar consignas de trabajo 3. CT= Concentrarse en la tarea 4. CC= Hacer cuadros comparativos 5. EO= Expresar por escrito u oralmente

				6. RE=. Resumir 7. RIP= Buscar y resumir ideas principales 8. LA=Leer atentamente 9. MR= Memorizar y repetir 10. RC= Establecen relaciones entre conceptos
45	estilo_aprendizaje	¿Qué estrategia utiliza para estudiar?	categoría	1. Visual_activo 2. Aural_reflexivo 3. Read_Lectura 4. Kinésico_pragmático
46	Auditivo	Estilos de enseñanza de los docentes	categoría	A
47	Lectura	Estilos de enseñanza de los docentes	categoría	R
48	Visual	Estilos de enseñanza de los docentes	categoría	V
49	Kinésico	Estilos de enseñanza de los docentes	categoría	K
50	Situación	Situación de estudio	categoría	{TERMINO, ABANDONO}

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

4.1.2.1. Análisis de los atributos

Los gráficos analizados en el presente apartado para el análisis de los atributos de la investigación son realizados por medio del software Weka el cual brinda una visión gráfica de cada una de las instancias utilizadas en el trabajo de investigación, permitiendo conocer el rango de valores que abarca cada instancia, su media, desviación estándar y la cantidad de valores perdidos que podrán dar una clara visión para realizar uno de los posteriores pasos como lo es la imputación de datos.

La decisión de utilizar Weka parte del análisis realizado en el marco referencial sobre dos tipos de software de preprocesado de datos, a continuación se detallan 35 de las 50 variables iniciales debido a que por decisión de la autora de la investigación se toman en cuenta las instancias de primer, segundo y tercer año de estudios.

Por medio del software seleccionado para el preprocesado de datos se reflejó que las instancias de los tres primeros años de estudio poseen un índice bajo de valores perdidos, además que el conjunto de reglas que se entreguen como resultado de la investigación están basadas sobre las mismas.

A continuación se estudian los gráficos obtenidos por medio del software Weka para el análisis de cada una de las instancias que contribuyeron en el desarrollo del trabajo, permitiendo considerar las técnicas que se deben aplicar para la obtención de mejores resultados.

4.1.2.1.1. Atributo: Carrera

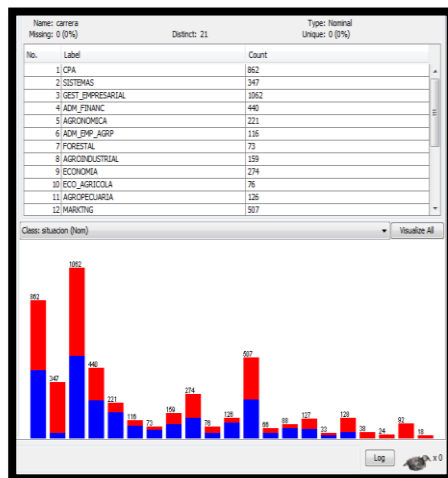


Ilustración 4.- Atributo de carrera

Interpretación de variable

En la presente gráfica se puede observar los datos de la instancia carrera en la cual se puede apreciar que la mayor parte de registros pertenecen a estudiantes de Ingeniería en Gestión empresarial, y un menor porcentaje en Ingeniería mecánica,

siendo así un total de 21 carreras a ser estudiadas en el trabajo investigativo, con un 0% de valores perdidos en el atributo lo que da un buen indicio a seguir con las etapas siguientes.

4.1.2.1.2. Atributo: Sexo



Ilustración 5.- Atributo Sexo

Interpretación de variable

En la siguiente gráfica se observa a la variable sexo de tipo nominal la cual posee dos características ya sea hombre o mujer, 2808 registros corresponden a mujeres frente a 2069 registros de hombres, en la instancia no se observa datos perdidos, dando como una descripción de la variable el género

de un estudiante permitiendo a la investigadora del trabajo contar con una variable sin valores perdidos.

4.1.2.1.3. Atributo: Promedio de primer año

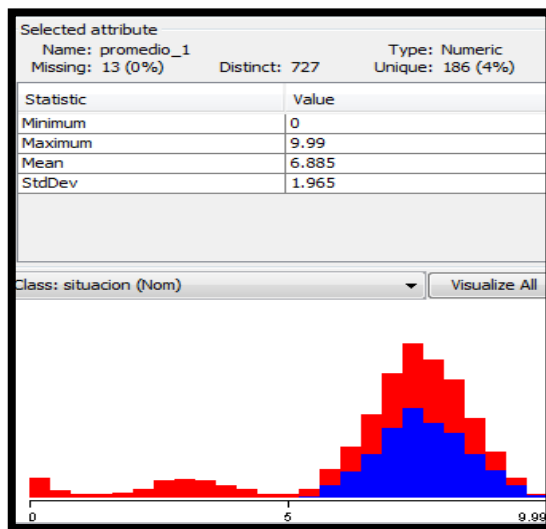


Ilustración 6.- Atributo primer año

Interpretación de variable

Se puede observar que en la presente variable se analiza los datos sobre el promedio de primer año de los estudiantes, la media de la población es de 6.88 y existen 13 valores perdidos los cuales pueden ser compensados en la etapa de imputación de datos, además tiene como valor máximo 9.99 y una desviación estándar de 1.96

4.1.2.1.4. Atributo Tiempo de aprobación de primer año

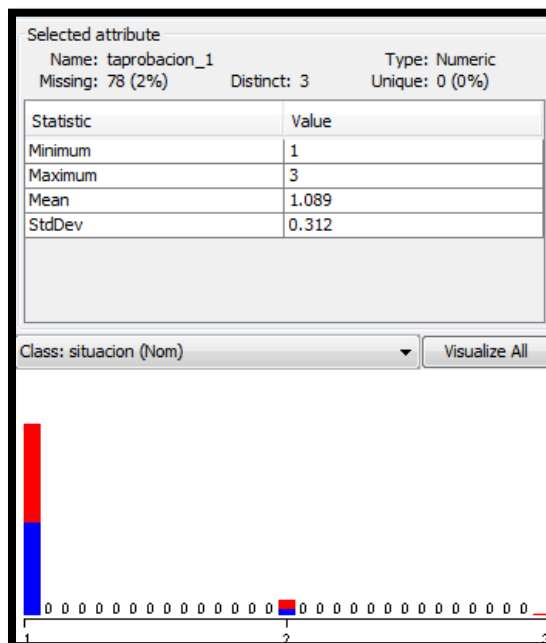


Ilustración 7.- Atributo tiempo de aprobación

Interpretación de variable

Mediante el análisis del presente gráfico se evidencia que la media del tiempo de aprobación de los estudiantes es de 1.08 con una desviación estándar 0.31 y que los datos poseen un 2% de valores perdidos lo cual sirvió para determinar la necesidad de usar técnicas de imputación, siendo el atributo que describa los años que tarda un estudiante para aprobar el mismo.

4.1.2.1.5. Atributo: Índice de aprobación de primer año

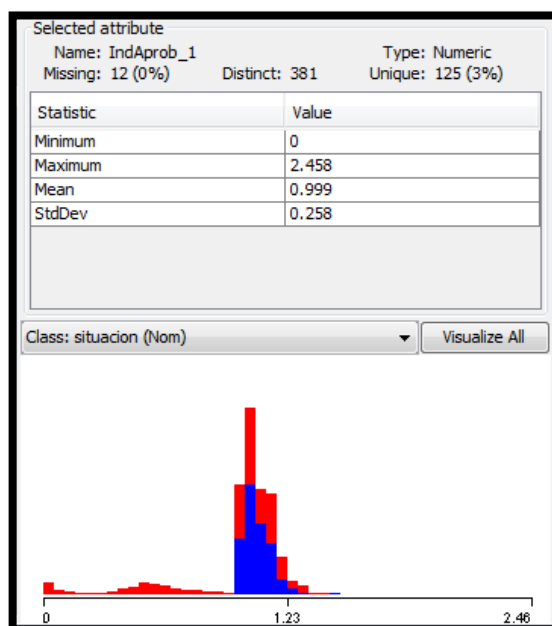


Ilustración 8.-Atributo Índice de aprobación de primer año

Interpretación de variable

En el siguiente gráfico se puede observar el índice de aprobación del primer año siendo la variable de tipo numérico la cual representa el número de materias aprobadas sin suspensión, además de analizar el valor de la media es de 0.99 con una desviación estándar de 0.25, y un 0% de valores perdidos lo que da pauta a seguir con el desarrollo de la investigación.

4.1.2.1.6. Atributo: Experiencia docente de primer año

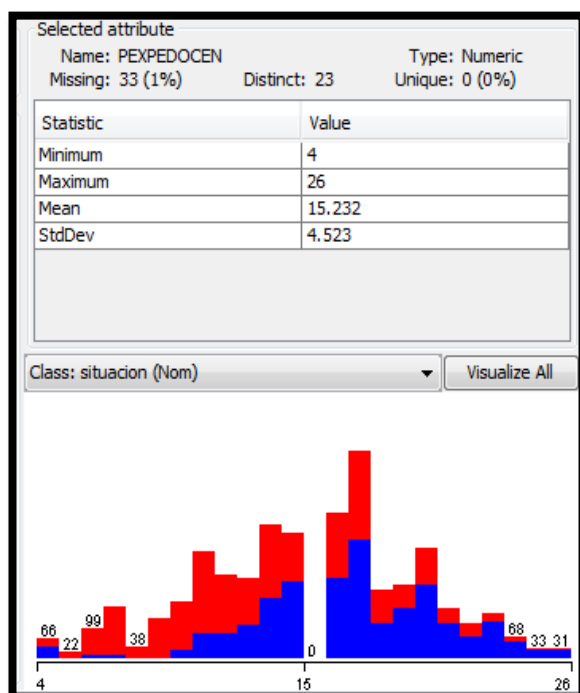
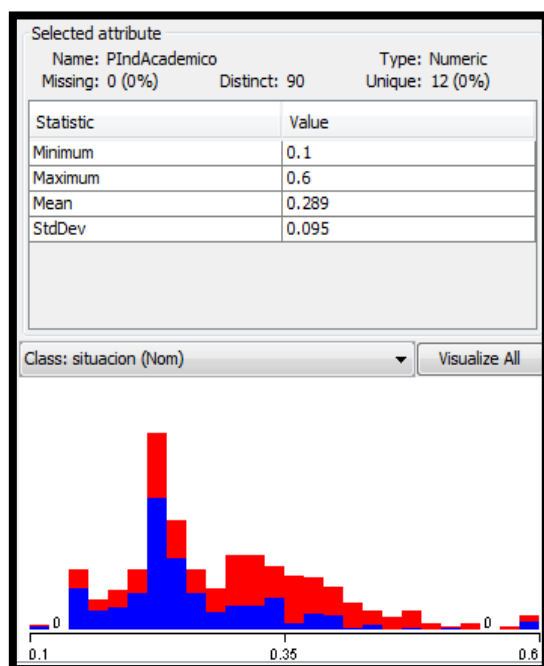


Ilustración 9.- Atributo experiencia docente de primer año

Interpretación de variable

Mediante el gráfico se puede observar los valores correspondientes a la instancia experiencia docente de primer año el cual está basado en el tiempo de enseñanza de los catedráticos, se evidencia en la gráfica una media de 15.23 con una desviación estándar 4.52 con un 1% de valores perdidos permitiendo desarrollar la investigación sin obviar el uso de algoritmos de imputación.

4.1.2.1.7. Atributo Índice académico de primer año



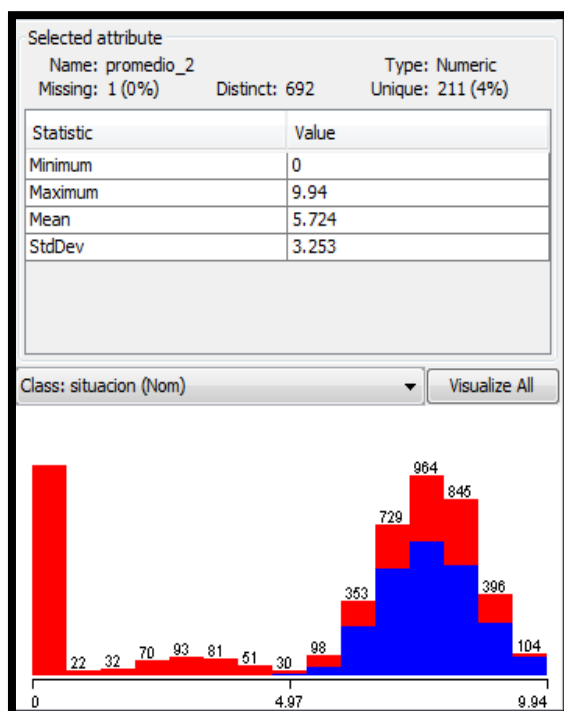
Interpretación de variable

En la presente ilustración se puede observar el índice académico de primer año el cual es el índice de aporte de la Universidad relacionada con la preparación académica del docente se puede observar los valores correspondientes a la media de 0.28 y de 0.09 con una desviación estándar 0.095 con un 0% de valores perdidos, permitiendo seguir con el

Ilustración 10.- Atributo Índice académico de primer año

desarrollo del trabajo investigativo.

4.1.2.1.8. Atributo: promedio de segundo año



Interpretación de variable

Se puede observar que la media del promedio de segundo año es de 5.72 y que hay 1 valor perdido equivaliendo a un 0% de la población total, teniendo como valor máximo 9.94 siendo estos los valores relevantes del promedio de segundo año.

Ilustración 11.- Atributo promedio de segundo año

4.1.2.1.9. Atributo: Tiempo de aprobación de segundo año

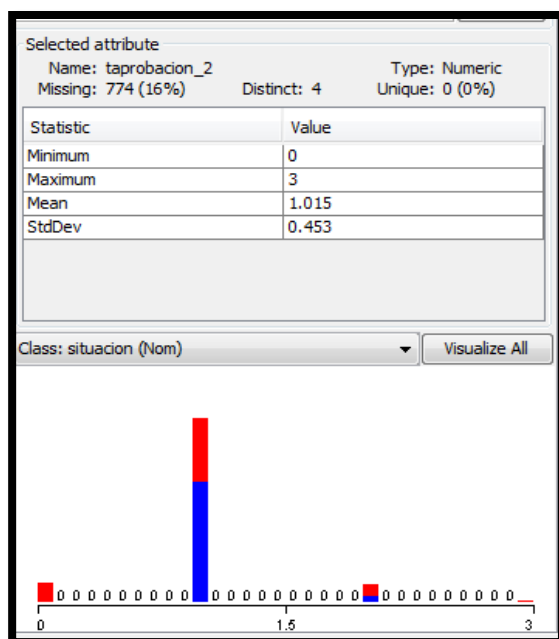


Ilustración 12 Atributo tiempo de aprobación de segundo año

Interpretación de variable

Mediante el análisis del presente gráfico se evidencia que la media del tiempo de aprobación de los estudiantes es de 1.015 con una desviación estándar 0.45 y que los datos poseen un 16% de valores perdidos en el atributo, siendo la instancia descrita la que represente los años que tarda un estudiante para aprobar el segundo año.

4.1.2.1.10. Atributo: Índice de aprobación de segundo año

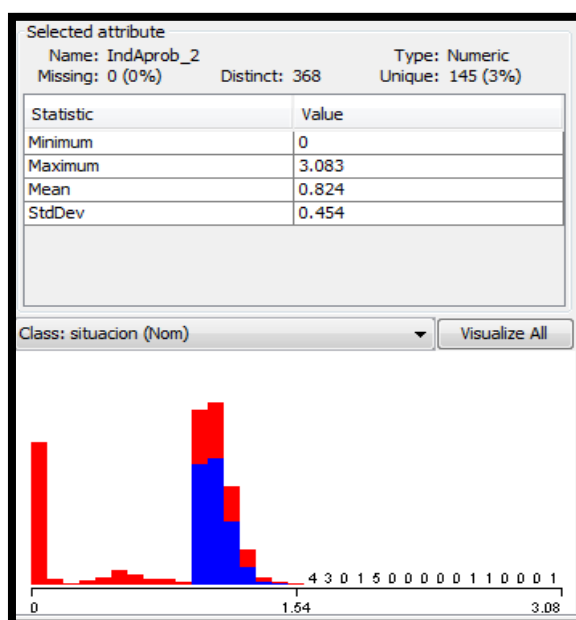


Ilustración 13 Atributo Índice de aprobación de segundo año

Interpretación de variable

En el siguiente gráfico se puede observar el índice de aprobación de segundo año siendo de tipo numérico, la cual representa el número de materias aprobadas sin suspensión, además de analizar el valor de la media es de 0.82 con una desviación estándar de 0.45, y un 0% de valores perdidos.

4.1.2.1.11. Atributo: Experiencia docente de segundo año

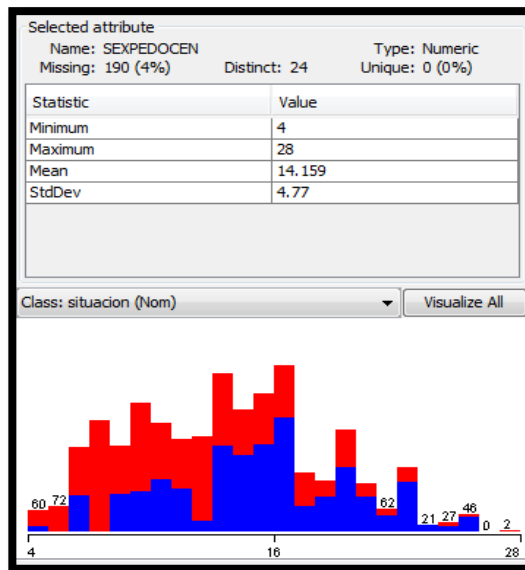


Ilustración 14 Experiencia docente de segundo año

Interpretación de variable

En el gráfico se puede observar los valores correspondientes a la instancia experiencia docente de segundo año el cual está basado en el tiempo de enseñanza de los catedráticos, se evidencia en la gráfica una media de 14.15 con una desviación estándar 4.77 con un 4% de valores perdidos valor que da indicio a utilizar los algoritmos de

imputación de datos.

4.1.2.1.12. Atributo Índice académico de segundo año

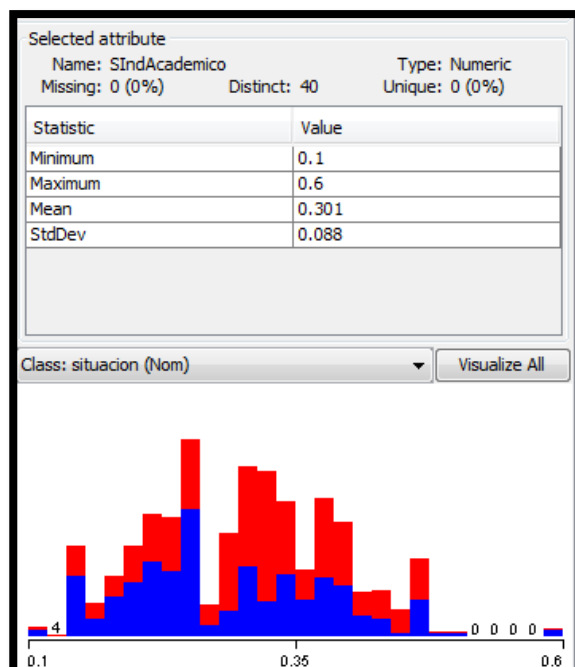


Ilustración 15 Atributo Índice académico de segundo año

Interpretación de variable

En la presente ilustración se puede observar el índice académico de segundo año el cual es el índice de aporte de la Universidad relacionada con la preparación académica del docente se puede observar los valores correspondientes a la media de 0.30 con una desviación estándar 0.088 con un 0% de valores perdidos.

4.1.2.1.13. Atributo: promedio de tercer año

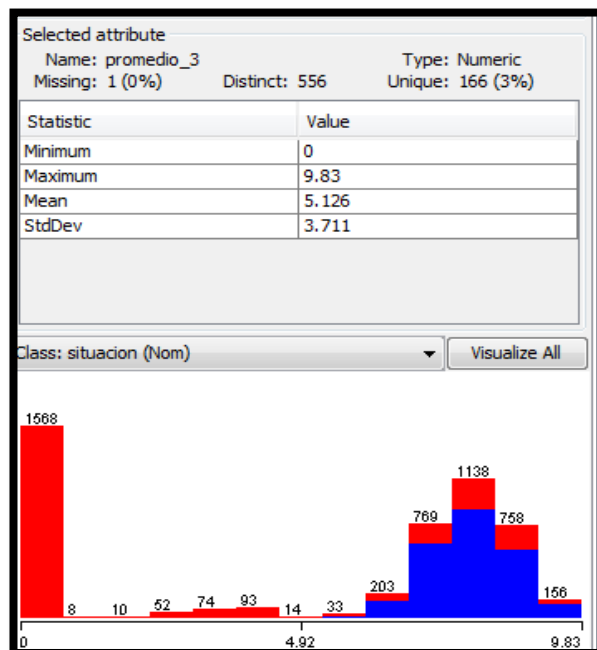


Ilustración 16 Atributo: promedio de tercer año

Interpretación de variable

La variable analizada corresponde al promedio de tercer año del conjunto de datos utilizados, se puede observar que la media de tercer año es de 5.12 y que hay 1 valor perdido equivaliendo a un 0% de la población total, teniendo como valor máximo 9.83 siendo estos los valores relevantes del promedio de tercer año.

4.1.2.1.14. Atributo: Tiempo de aprobación de tercer año

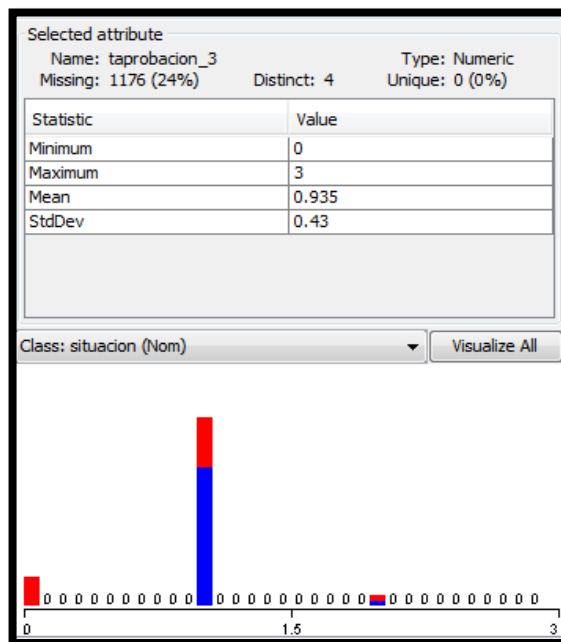


Ilustración 17 Atributo: Tiempo de aprobación de tercer año

Interpretación de variable

Mediante el análisis del presente gráfico se evidencia que la media del tiempo de aprobación de los estudiantes es de 0.93 con una desviación estándar 0.43 y que los datos poseen un 24% de valores perdidos en el atributo, siendo el mismo el que describe los años que tarda un estudiante para aprobar el año.

4.1.2.1.15. Atributo: Índice de aprobación de tercer año

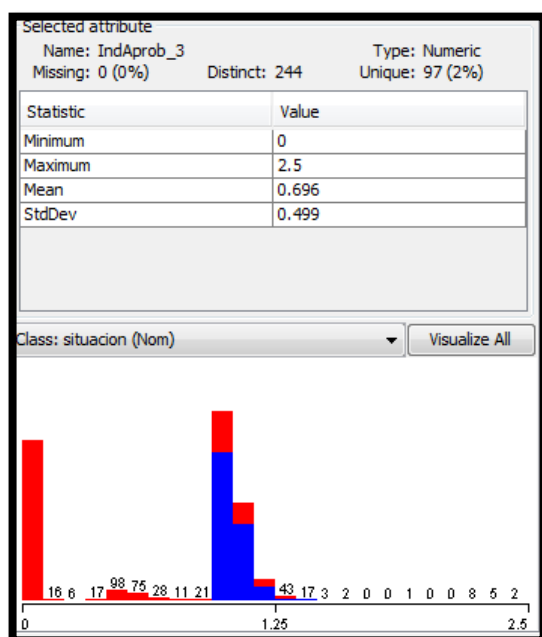


Ilustración 18 Índice de aprobación de tercer año

Interpretación de variable

En el siguiente gráfico se puede observar el índice de aprobación de tercer año siendo la variable analizada de tipo numérica la cual representa el número de materias aprobadas sin suspensión, además de analizar el valor de la media es de 0.69 con una desviación estándar de 0.49, y un 0% de valores perdidos.

4.1.2.1.16. Atributo Experiencia docente de tercer año

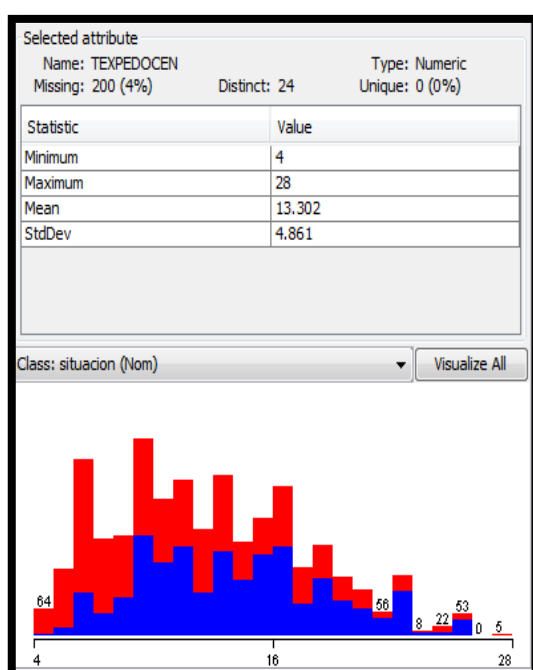


Ilustración 19 Atributo experiencia docente de tercer año

Interpretación de variable

En el gráfico se puede observar los valores correspondientes a la instancia experiencia docente de tercer año el cual se basa en el tiempo de enseñanza de los catedráticos, se evidencia en la gráfica una media de 13.30 con una desviación estándar 4.86 con un 4% de valores perdidos, generando la necesidad de usar métodos de imputación.

4.1.2.1.17. Atributo índice académico de tercer año

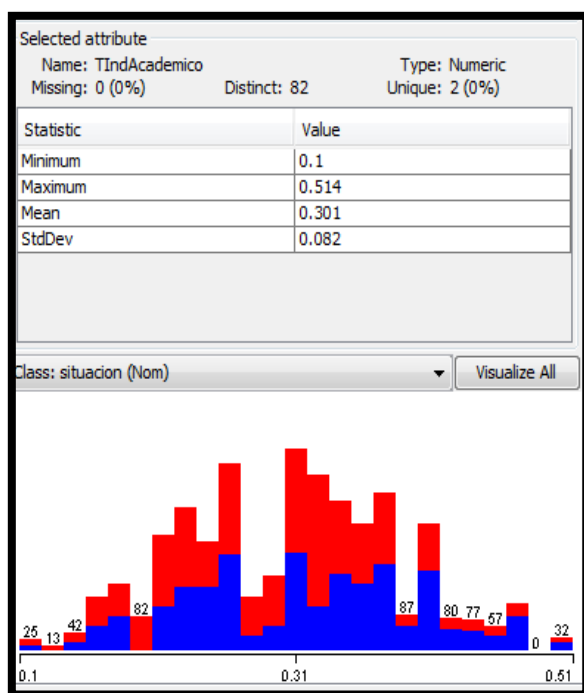


Ilustración 20 Atributo índice académico de tercer año

Interpretación de variable

En la presente ilustración se puede observar el índice académico de tercer año el cual es el índice de aporte de la Universidad relacionada con la preparación académica del docente, se puede observar los valores correspondientes a la media de 0.30 con una desviación estándar 0.082 con un 0% de valores perdidos, dando así paso a las siguientes etapas del desarrollo del

trabajo.

4.1.2.1.18. Atributo: Promedio de cuarto año

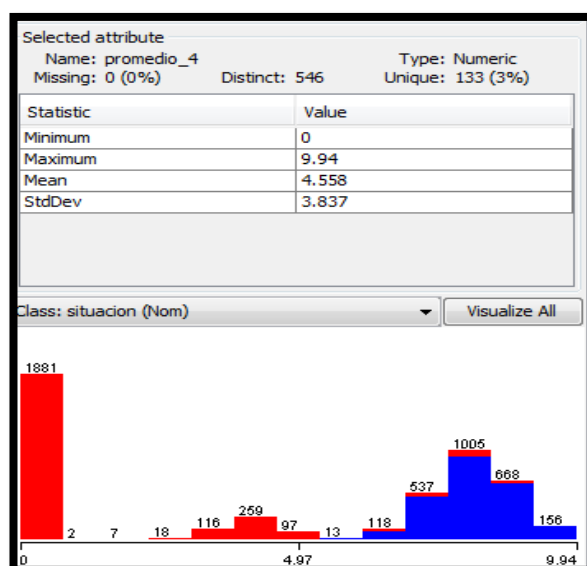


Ilustración 21 Atributo promedio de cuarto año

Interpretación de variable

La presente instancia representa el promedio de cuarto año de los estudiantes, se puede observar que la media de cuarto año es de 6.88 y que hay 13 valores perdidos equivaliendo a un 0% de la población total, teniendo como valor máximo 9.99, dando lugar al desarrollo de la siguiente etapa de la investigación.

4.1.2.1.19. Atributo Tiempo de aprobación de cuarto año

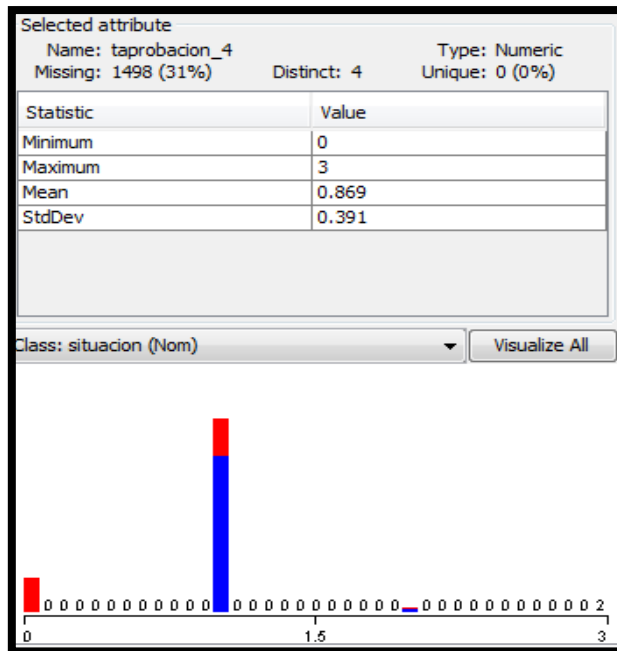


Ilustración 22 Atributo tiempo de aprobación de cuarto año

necesidad utilizar métodos de imputación.

Interpretación de variable

Mediante el análisis del presente gráfico se evidencia que la media del tiempo de aprobación de los estudiantes es de 0.86 con una desviación estándar 0.39 y que los datos poseen un 31% de valores perdidos, siendo el presente atributo el que describa los años que tarda un estudiante para aprobar el mismo, generando así la

4.1.2.1.20. Atributo: Índice de aprobación de cuarto año

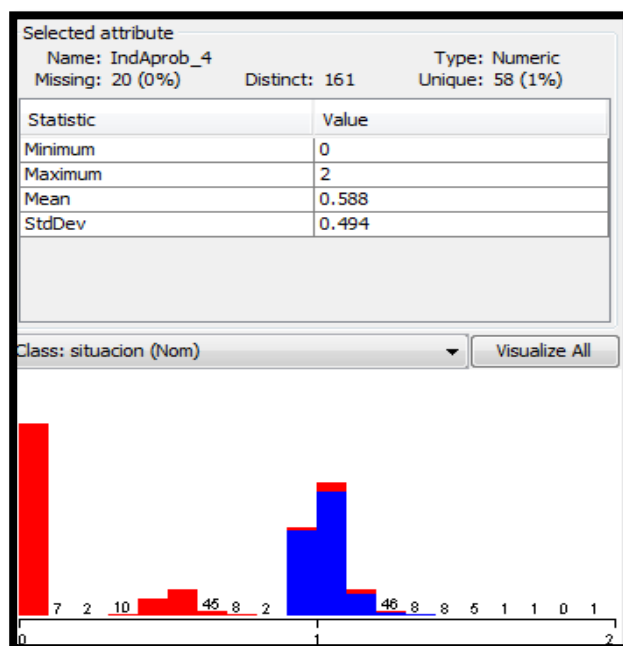


Ilustración 23 Atributo Índice de aprobación de cuarto año

Interpretación de variable

En el siguiente gráfico se puede observar el índice de aprobación del cuarto año siendo la variable de tipo numérico la cual representa el número de materias aprobadas sin suspensión, además de analizar el valor de la media es de 0.58 con una desviación estándar de 0.49, y un 0% de valores perdidos.

4.1.2.1.21. Atributo: Experiencia docente de cuarto año

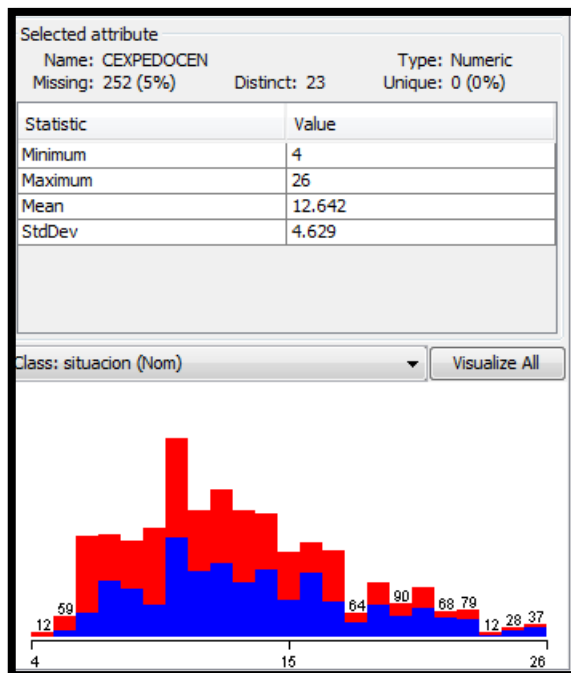


Ilustración 24 Atributo experiencia docente de cuarto año

Interpretación de variable

En el gráfico se puede observar los valores correspondientes a la instancia experiencia docente de cuarto año el cual está basado en el tiempo de enseñanza de los catedráticos, se evidencia en la gráfica una media de 12.64 con una desviación estándar 4.62 con un 5% de valores perdidos, lo cual permite tomar la decisión de usar métodos de imputación.

4.1.2.1.22. Atributo Índice académico de cuarto año

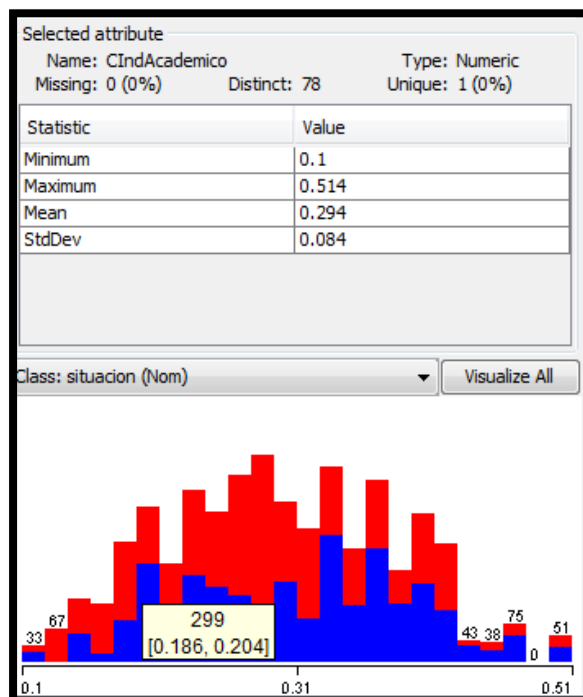


Ilustración 25 Atributo Índice académico de cuarto año

Interpretación de variable

En la presente ilustración se puede observar el índice académico de cuarto año el cual es el índice de aporte de la Universidad relacionada con la preparación académica del docente se puede evidenciar los valores correspondientes a la media de 0.29 con una desviación estándar 0.084 con un 0% de valores perdidos.

4.1.2.1.23. Atributo: promedio de quinto año

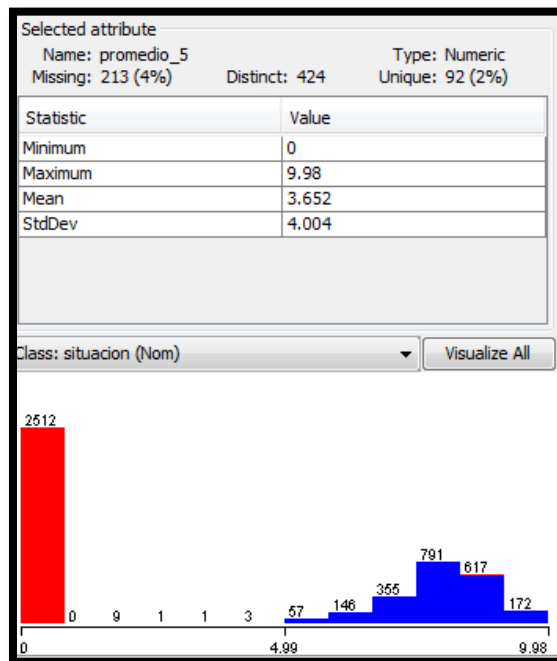


Ilustración 26 Atributo promedio de quinto año

Interpretación de variable

Se puede observar que la media del promedio de quinto año de los estudiantes es de 3.65 y que hay 213 valores perdidos equivaliendo a un 4% de la población total, teniendo como valor máximo 9.98 siendo estos los valores relevantes del promedio de quinto año, y generando la necesidad de

aplicar la imputación de datos.

4.1.2.1.24. Atributo: Tiempo de aprobación de quinto año

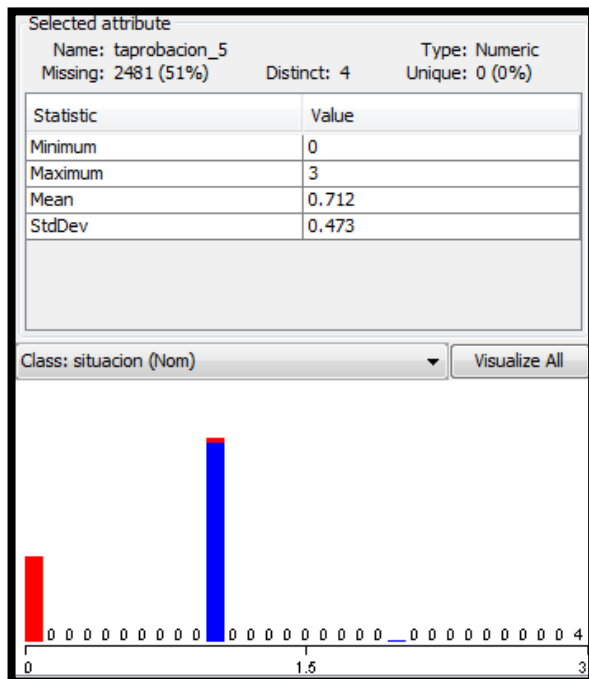


Ilustración 27 Atributo tiempo de aprobación de quinto año

Interpretación de variable

Mediante el análisis del presente gráfico se evidencia que la media del tiempo de aprobación de los estudiantes es de 0.71 con una desviación estándar 0.47 y que los datos poseen un 51% de valores perdidos en el atributo, siendo el mismo el que describa los años que tarda un estudiante para aprobar el quinto año.

4.1.2.1.25. Atributo: Índice de aprobación de quinto año

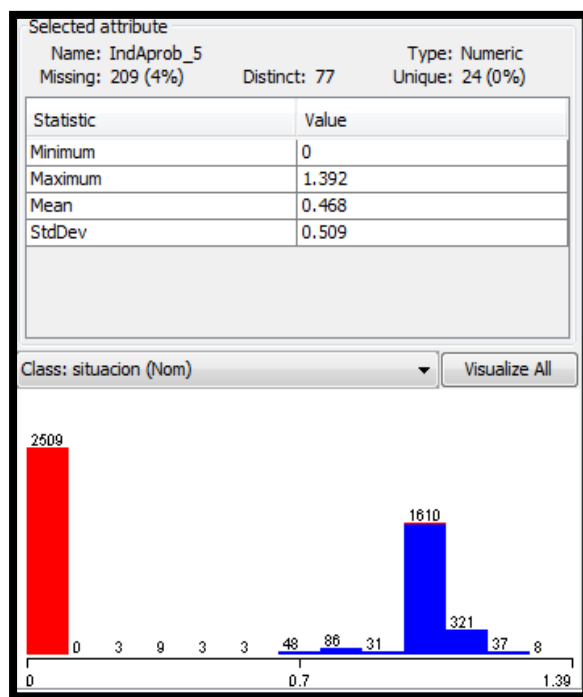


Ilustración 28 Atributo Índice de aprobación de quinto año

Interpretación de variable

En el siguiente gráfico se puede observar el índice de aprobación de quinto año siendo la variable de tipo numérico la cual representa el número de materias aprobadas sin suspensión, además de analizar el valor de la media es de 0.46 con una desviación estándar de 0.50, y un 4% de valores perdidos, valores que podrían ser compensados con la imputación de datos.

4.1.2.1.26. Atributo: Experiencia docente de quinto año

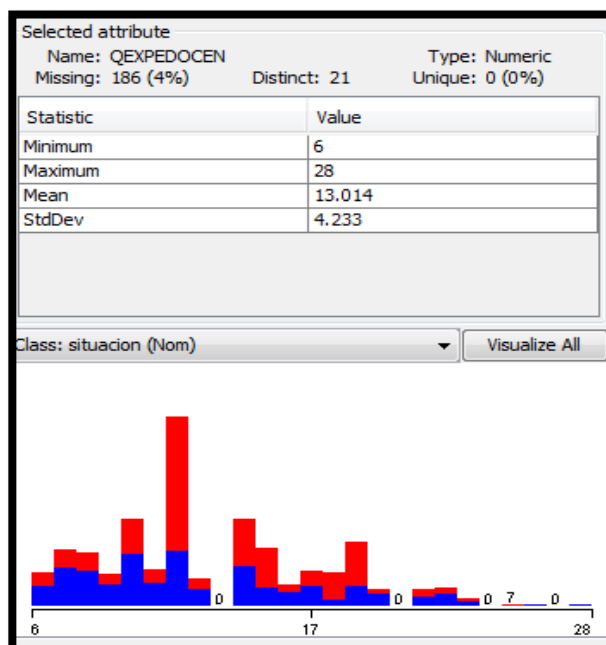


Ilustración 29 Experiencia docente de quinto año

Interpretación de variable

En el gráfico se puede observar los valores correspondientes a la instancia experiencia docente de quinto año el cual está basado en el tiempo de enseñanza de los catedráticos, se evidencia en la gráfica una media de 13.01 con una desviación estándar 4.23 con un 4% de valores perdidos.

4.1.2.1.27. Atributo Índice académico de quinto año

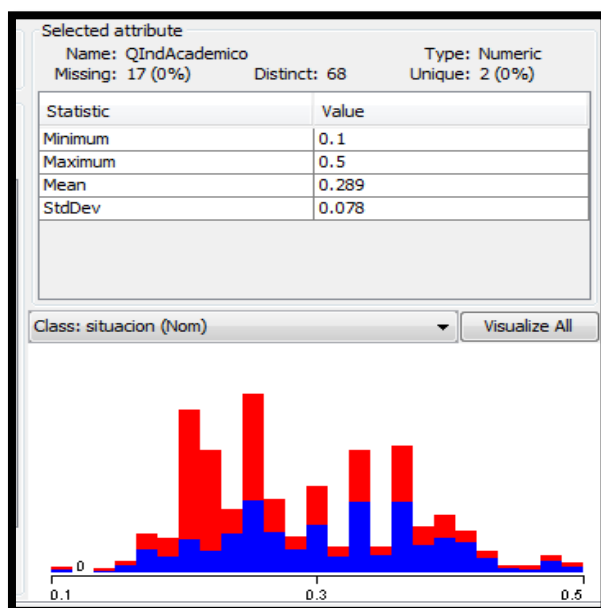


Ilustración 30 Atributo Índice académico de quinto año

Interpretación de variable

En la presente ilustración se puede observar el índice académico de quinto año el cual es el índice de aporte de la Universidad relacionada con la preparación académica del docente se puede observar los valores correspondientes a la media de 0.28 con una desviación estándar 0.078 con un 0% de valores perdidos.

4.1.2.1.28. Atributo: habito de estudio

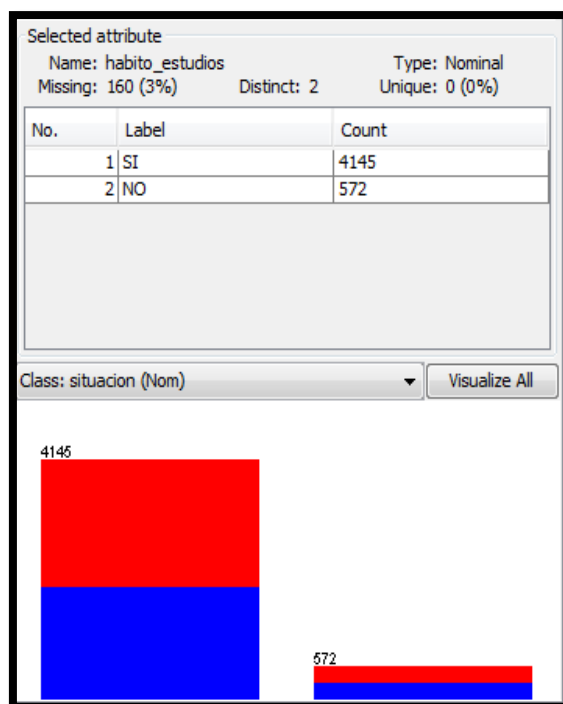
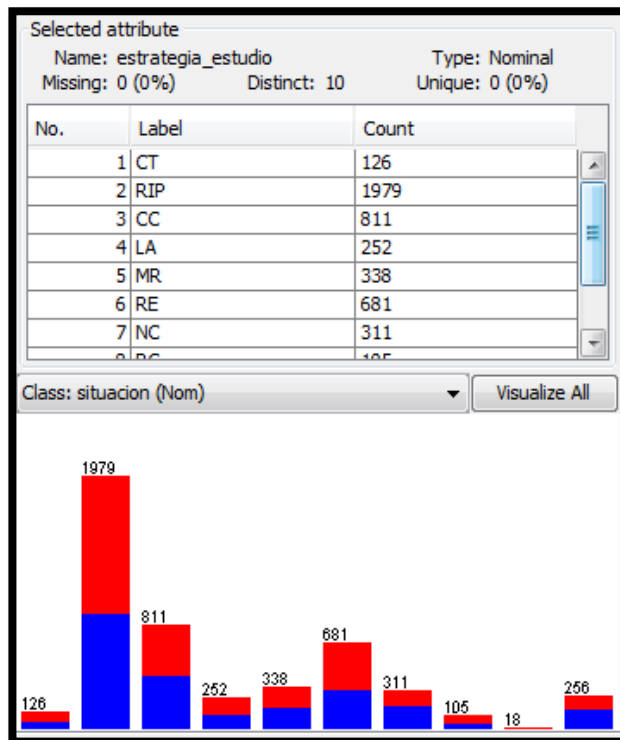


Ilustración 31 Atributo hábitos de estudio

Interpretación de variable

Mediante el gráfico presentado se puede deducir que la mayor parte de estudiantes de la Universidad Técnica Estatal de Quevedo poseen hábitos de estudios, además se puede observar que existe un 3 % de datos perdidos lo que permite realizar en la siguiente etapa la imputación de datos.

4.1.2.1.29. Atributo: Estrategia de estudio

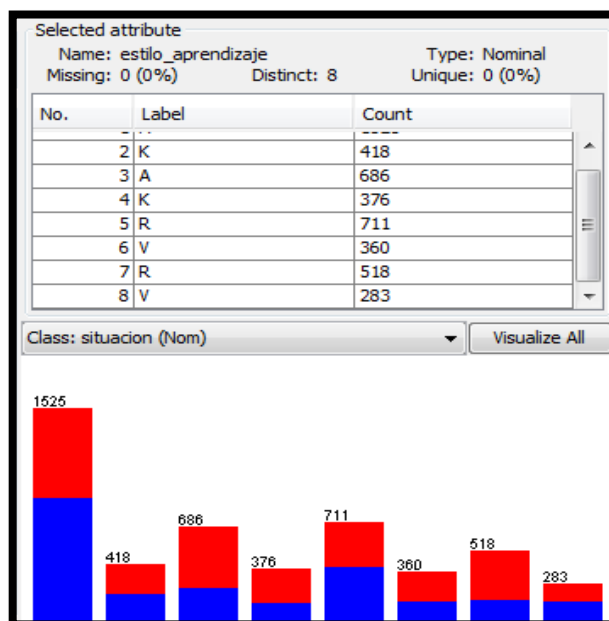


Interpretación de variable

Se puede analizar según el gráfico estadístico que la estrategia de estudio Resumir ideas principales es una de las más usadas en la universidad frente a Interpretar consignas de trabajo que es la de menor porcentaje, además se puede observar que no existen valores perdidos.

Ilustración 32 Estrategia de estudio

4.1.2.1.30. Atributo: Estilo de aprendizaje



Interpretación de variable

Se puede observar que de los estilos de aprendizajes VARK la mayor parte de estudiantes utilizan el estilo de aprendizaje auditivo y la menor parte corresponden a visual, no se encuentran valores perdidos

Ilustración 33 Atributos Estilo de aprendizaje

4.1.2.1.31. Atributo: Auditivo

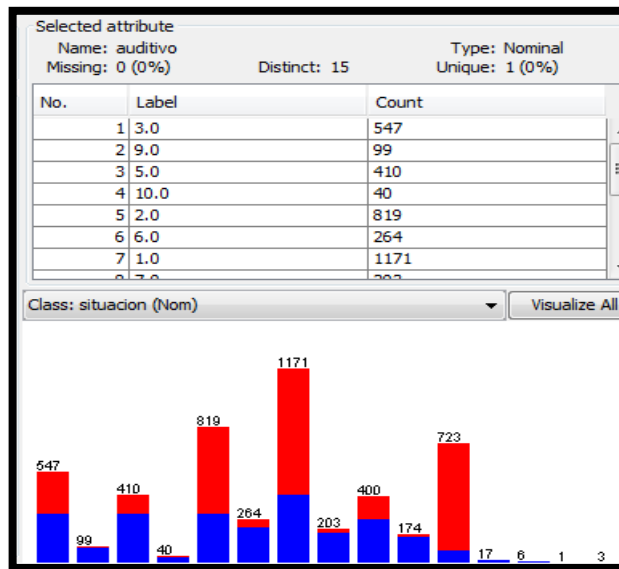


Ilustración 34 Atributo auditivo

Interpretación de variable

La instancia Auditivo es estilo de enseñanza que refleja las preferencias de escuchar información de ciertos estudiantes, en el presente gráfico se puede observar que la mayor parte de registros tienen al menos un docente con estilo de enseñanza auditivo, y que no existen valores perdidos

4.1.2.1.32. Atributo: Lectura

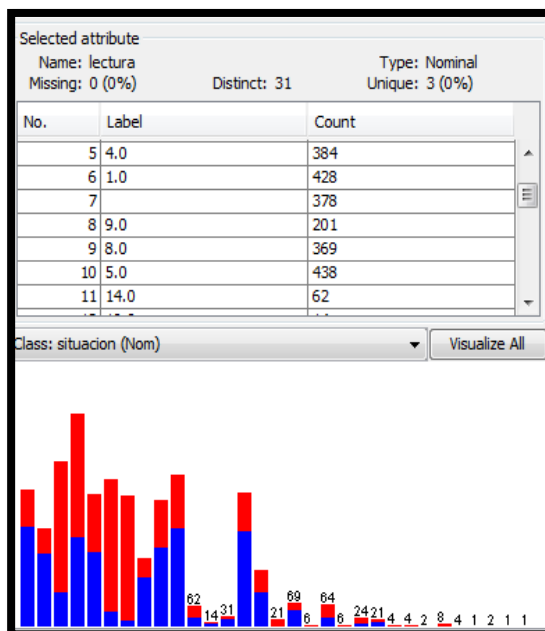


Ilustración 35 Atributo Lectura

Interpretación de variable

La variable Lectura expresa la preferencia por la información impresa en forma de palabras, según lo observado en los gráficos se puede deducir que una gran parte de docentes imparten clases lo realizan por medio del estilo de enseñanza de lectura, se puede ver que no existen valores perdidos

4.1.2.1.33. Atributo: Visual

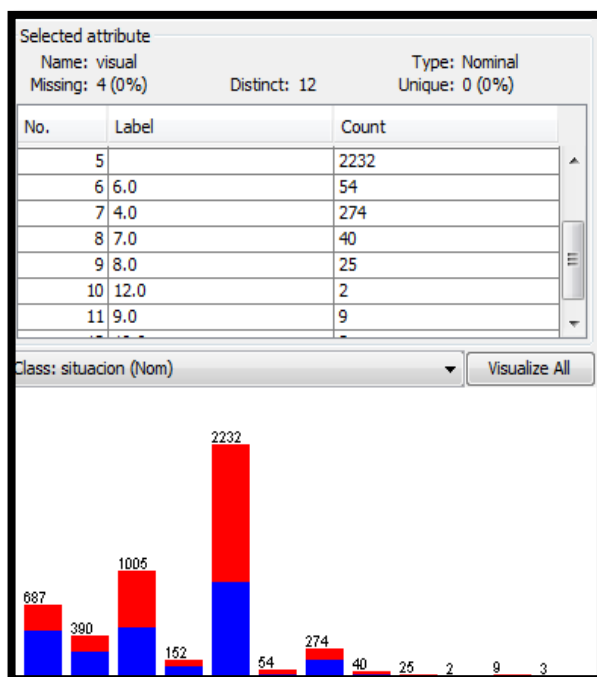


Ilustración 36 Atributo Visual

Interpretación de variable

La instancia Visual refleja la preferencia por maneras gráficas y simbólicas de representar la información, el estilo de enseñanza visual se inclina en que la mayor parte de veces al menos 1 docente imparte clases utilizando el estilo existe pocos valores perdidos.

4.1.2.1.34. Atributo: Kinésico

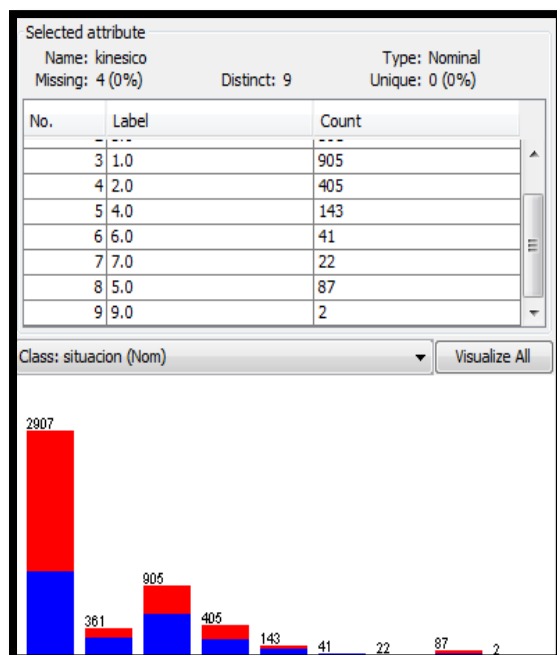


Ilustración 37 Atributo Kinésico

Interpretación de variable

El atributo Kinésico se basa en la preferencia de la experiencia y la práctica, el presente estilo de enseñanza kinésico se inclina en que la mayor parte de veces al menos 1 docente imparte clases utilizando el estilo existe pocos valores perdidos dando un 0%.

4.1.2.1.35. Atributo: Situación

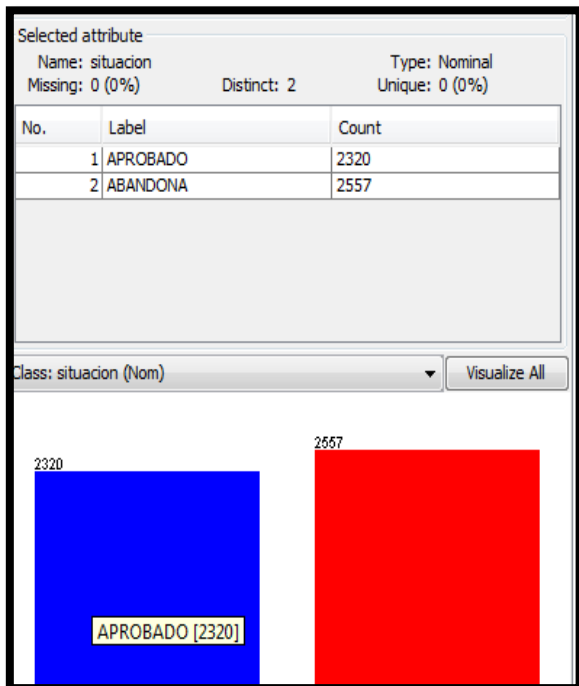


Ilustración 38 Atributo situación

Interpretación de variable

En la siguiente gráfica se observa la clase del conjunto de datos que servirán para la extracción de conocimiento comenzado como base con dos situaciones: aprobado, abandona; se puede observar que se cuenta con mayor número de registros de la clase aprobado y que no existen valores perdido correspondiendo estos a un 0%.

4.1.2.2. Transformación y selección de atributos

En el desarrollo de la tesis se elaboró la selección de atributos en base a instancias que aporten a las características de la investigación, luego se procedió a la transformación de los mismos para que estén en un formato que puedan ser procesables en los métodos de Minería de datos.

En el trabajo de investigación se inició con 50 instancias que cumplían con las características generales del problema estudiado, la investigadora del proyecto opto por trabajar con 35 atributos debido a que los atributos de cuarto año en adelante no serían muy útil si se desea predecir la tendencia de aprobación o reprobación de un alumno desde los primeros años de estudios.

En el trascurso de la elaboración de la extracción de conocimiento entre los estilos de enseñanza aprendizaje se logró disminuir las variables a un total de 13 siendo estas las que aportan más al modelo y que poseen características más relevantes.

Para la selección de atributos se utilizó el software Weka el cual permitió a la autora de la tesis analizar los atributos en base a los gráficos que muestra, a la desviación estándar, a los máximos, mínimos y valores nulos.

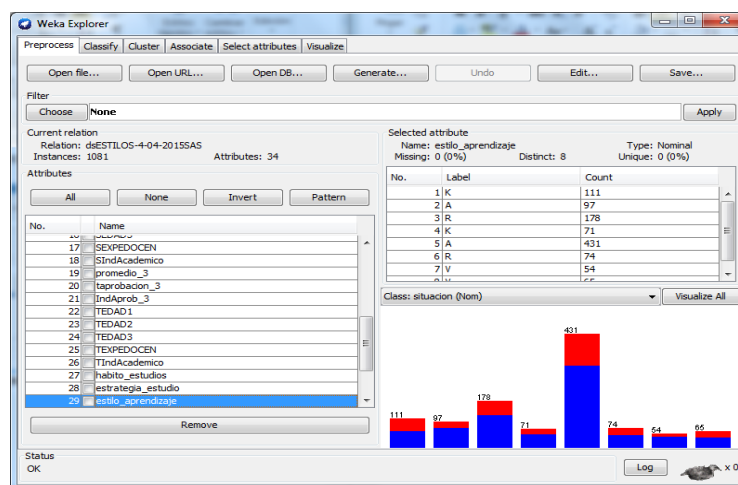


Ilustración 39.- Análisis de atributos en weka

Por medio de la utilización del software se pudo aplicar la selección de atributos con diferentes algoritmos de búsqueda de atributos con mayor índice de aportación al modelo, lo que permitió formar grupos de instancias como se muestra en la Ilustración 40 en la que se observa la selección de atributos dando lugar a una vista minable.

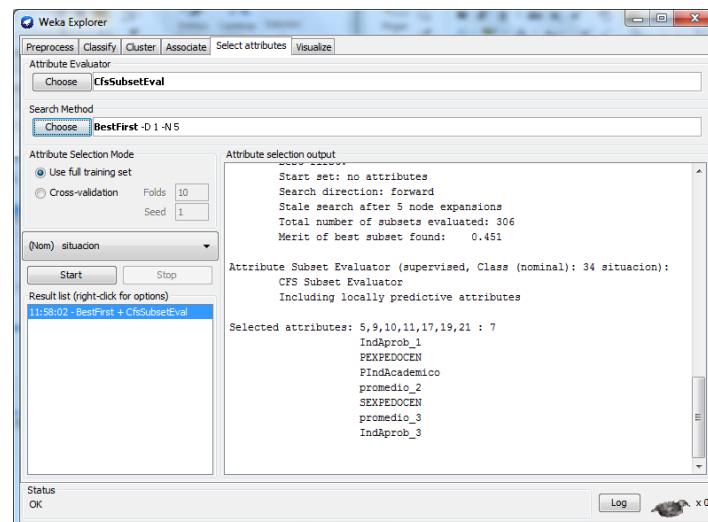


Ilustración 40.- Selección de atributos en weka

4.1.2.3. Transformación de atributos

La etapa de transformación de atributo consiste en convertir los datos en un formato que sea procesable por el método de minería de dato. Para lo cual la autora del trabajo de investigación realizó la transformación del atributo Edad_docente por medio de rangos que permitan categorizar a la edad de cada uno de los docentes en atributos tales como: PEDAD1, PEDAD2, PEDAD3 siendo Rango de edad de docentes menores a 35 años, Rango de edad de docentes entre 35 y 50 años, Rango de edad de docentes mayores a 50 años respectivamente y así para cada uno de los años de estudio, lo cual permitió un mejor análisis de la situación.

4.1.2.4. Vistas minables

Con la utilización del software Weka y los algoritmos de búsqueda como: Bestfirst, GenethicSearch, entre otros se logró realizar una parte del preprocesado de datos como lo es la selección de atributos, por lo cual a continuación se detallan los conjuntos de instancias obtenidas para la elaboración del proyecto, descritas como vistas minables:

Tabla 8.- Vista minable N° 1

NÚMERO DE VISTAS MINABLES	ATRIBUTOS SELECCIONADOS
VISTA MINABLE N° 1	Promedio_2
	SEXPEDOCEN
	Promedio_3
	TIndiceAcademico
	Promedio_4
	IndAprob_4
	CEXPEDOCEN
	Promedio_5
	IndiceAprob_5
	QEDAD2
	QEDAD3
	Habito_estudio
	Estilo_aprendizaje
	Auditivo
	Lectura
	Visual
	Kinésico

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

Tabla 9.- Vista Minable N° 2

NÚMERO DE VISTAS MINABLES	ATRIBUTOS SELECCIONADOS
VISTA MINABLE N° 2	PIndAcademico
	promedio_2
	taprobacion_2
	IndAprob_2
	promedio_3
	IndAprob_3
	Habito_estudio
	Estilo_aprendizaje
	Auditivo
	Lectura
	Visual
	Kinésico

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

Tabla 10.- Vista minable N° 3

NÚMERO DE VISTAS MINABLES	ATRIBUTOS SELECCIONADOS
VISTA MINABLE N° 3	IndAprob_1
	PEXPEDOCEN
	PIndAcademico
	promedio_2
	SEXPEDOCEN
	promedio_3
	IndAprob_3
	Habito_estudio
	Estilo_aprendizaje

	Auditivo
	Lectura
	Visual
	Kinésico

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

Tabla 11.- Vista minable N° 4

NÚMERO DE VISTAS MINABLES	ATRIBUTOS SELECCIONADOS
VISTA MINABLE N° 4	IndAprob_3
	promedio_3
	PEDAD1
	TEDAD2
	SEXPEDOCEN
	SEDAD1
	Carrera
	PIndAcademico
	PEXPEDOCEN
	TIndAcademico
	IndAprob_2
	promedio_2
	Estilo_aprendizaje
	Auditivo
	Lectura
	Visual
	Kinésico

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

Tabla 12.- Vista minable N° 5

NÚMERO DE VISTAS MINABLES	ATRIBUTOS SELECCIONADOS
VISTA MINABLE N° 5	IndAprob_3
	PEDAD1
	TEDAD2
	SEDAD1
	Carrera
	PIndAcademico
	PEXPEDOCEN
	TIndAcademico
	IndAprob_2
	Estilo_aprendizaje
	Auditivo
	Lectura
	Visual

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

4.1.2.5. Imputación de datos

En el proceso de extracción de conocimiento es muy común encontrar datos incompletos lo cual resulta complejo al momento de procesar los mismos, en el caso de estudio se encontró inconsistencia en los datos debido a diversos factores ya sea por: personas que no respondieron las encuestas, error al guardar los datos.

En el análisis de los atributos realizado anteriormente se pudo observar como existen variables que tienen porcentaje altos de valores perdidos, por lo cual se genera la necesidad de utilizar algoritmos de imputación de datos.

Para la imputación de datos en la extracción de conocimiento mediante sistemas difusos de los estilos de enseñanza/ aprendizaje entre los docentes y estudiantes de la UTEQ se utilizó el software Keel, debido a que presenta una interfaz amigable , cuenta con diversos algoritmos de imputación y permite una fácil comprensión de los resultados.

El software Keel brinda a los investigadores diversos algoritmo de imputación de datos, para el actual estudio se utilizó el algoritmo *Fuzzy K-means Imputation* como se muestra en la **Ilustración 41**, el cual trabaja dados k intervalos intentando minimizar la distancia entre cada valor y el valor medio de su intervalo correspondiente, logrando completar los valores perdidos de los atributos analizados.

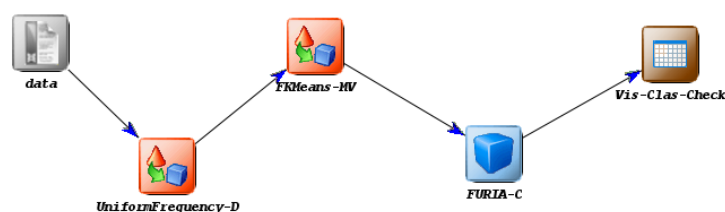


Ilustración 41.- Imputación de datos FKmeans

4.1.2.6. Técnicas de discretización

En la elaboración del trabajo se discretizó manualmente varios atributos que estaban de forma continua a categórica, además de aplicar los algoritmos que se encuentran en el software Keel como se muestra en la **Ilustración 42**, se puede observar que el nodo aplicado para la discretización en el trabajo de investigación fue *Uniform Frequency Discretizer*, el cual se utiliza para dividir en grupos los atributos de tal forma que se realicen menor número de pruebas.

La aplicación de algoritmos de discretización se la realizó para reducir los tiempos de ejecución de un proyecto, además de que permitió disminuir el modelo resultante es decir se tuvo menor cantidad de reglas y se logró facilitar la interpretación de las mismas.

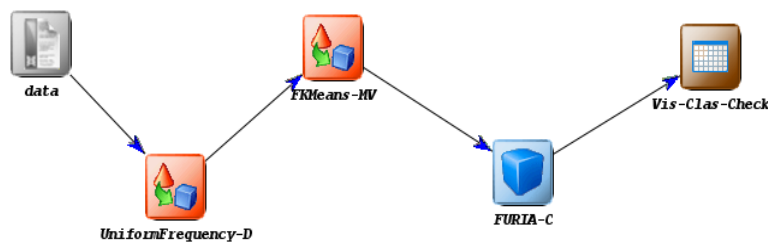


Ilustración 42.- Discretizar datos en Keel

4.1.3. Extracción de conocimiento

En la etapa de la extracción de conocimiento se aplicarán diversas técnicas difusas que permitirán modelar reglas que definan la relación de los diferentes tipos de enseñanza /aprendizaje entre docentes y estudiantes de la Universidad Técnica Estatal de Quevedo

La aplicación de algoritmos difusos no garantiza encontrar buenos patrones, en la aplicación de diversas técnicas difusas pueden observarse resultados que poseen índices de precisión altos pero a su vez también otros que tienen muy

bajos resultados, probablemente siendo esto último por factores como ruido en la información o porque no se encuentre un patrón entre los datos.

A continuación se detallan los algoritmos difusos utilizados en la investigación con los índices de prueba y entrenamiento hechos en cada experimento con la utilización del software keel.

4.1.3.1. Análisis de resultados de algoritmos

En la elaboración del trabajo de investigación se trabajó con 5 vistas minables a las cuales se le aplicaron varios algoritmos difusos pero debido a que los índices de precisión oscilaban entre 60 -70 % se descartaron cuatro vistas minables, siendo la **vistas minable N°2** la seleccionada para realizar pruebas con nueve algoritmos difusos que se encuentran en Keel con la imputación de datos *Fuzzy K-means Imputation*, a continuación en la **Tabla 13** se podrá observar la descripción de los algoritmos aplicados en la investigación.

Tabla 13.-Descripción de los algoritmos de los experimentos

ITEM	ALGORITMO	TIPO	ENFOQUE	REFERENCIA
1	FURIA-C	Clasificación	Fuzzy Rule Learning	(Huhn & Hullermeier., 2009)
2	Hybrid Fuzzy GBML (FH-GBML-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Ishibuchi, Yamamoto & Nakashima, 2005)
3	AdaBoost algorithm (AdaBoost-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Del Jesús, Hoffmann, Junco & Sánchez,2004)
4	Grammar-based GP algorithm (GFS-GP-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Sánchez & Couso, 2001)

5	Grammar-based GAP algorithm (GFS-GPG-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Sánchez & Couso, 2001)
6	LogitBoost algorithm (GFS-LogitBoost-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Otero & Sánchez, 2006)
7	Fuzzy Rule Learning with Single Winner Inference, Logitboost algorithm (GFS-MaxLogitBoost-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Otero & Sánchez, 2007)
8	Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules From Data (SGERD-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Mansoori,Zolghadri & Katebi, 2008)
9	Iterative Rule Learning of Fuzzy Rules with Feature Selection (SLAVE-C)	Clasificación	Evolutionary Fuzzy Rule Learning	(Gonzalez & Perez, 2001)

AUTORA: Olvera Toala Lely Johanna

FUENTE: Investigación

4.1.3.2. Resultados obtenidos con la aplicación de los algoritmos difusos para la comprobación de la hipótesis planteada

A continuación se muestran los resultados logrados con las técnicas difusas mediante el uso del software Keel, para lo cual se muestran gráficos que permiten una mejor visión de los mismos y que contribuyen a la comprobación de la hipótesis planteada:

4.1.3.2.1. Análisis del conjunto de reglas obtenidas

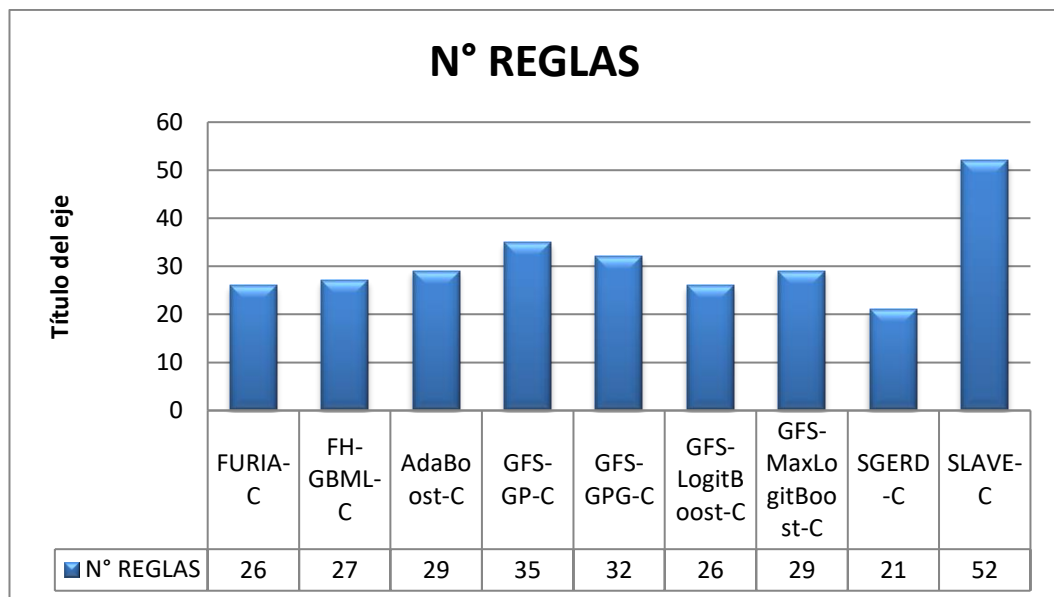


Ilustración 43.- Análisis de reglas

Según la Ilustración 43 se puede observar que el algoritmo que devuelve menor número de reglas es el Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules From Data (SGERD-C) frente al Iterative Rule Learning of Fuzzy Rules with Feature Selection (SLAVE-C) que devuelve 52 reglas.

Con lo anteriormente observado la autora de la investigación deduce que según el análisis de la ilustración anterior el mejor resultado es el del algoritmo SGRED-C con 21 reglas, cabe recalcar que para el análisis no se tomaron en cuenta la calidad de la reglas.

Tabla 14 Análisis del conjunto de reglas de algoritmos difusos

ITEM	ALGORITMO	N° REGLAS
1	FURIA-C	26
2	FH-GBML-C	27
3	AdaBoost-C	29

4	GFS-GP-C	35
5	GFS-GPG-C	32
6	GFS-LogitBoost-C	26
7	GFS-MaxLogitBoost-C	29
8	SGERD-C	21
9	SLAVE-C	52

4.1.3.2.2. Análisis del conjunto de la desviación estándar de los resultados obtenidos

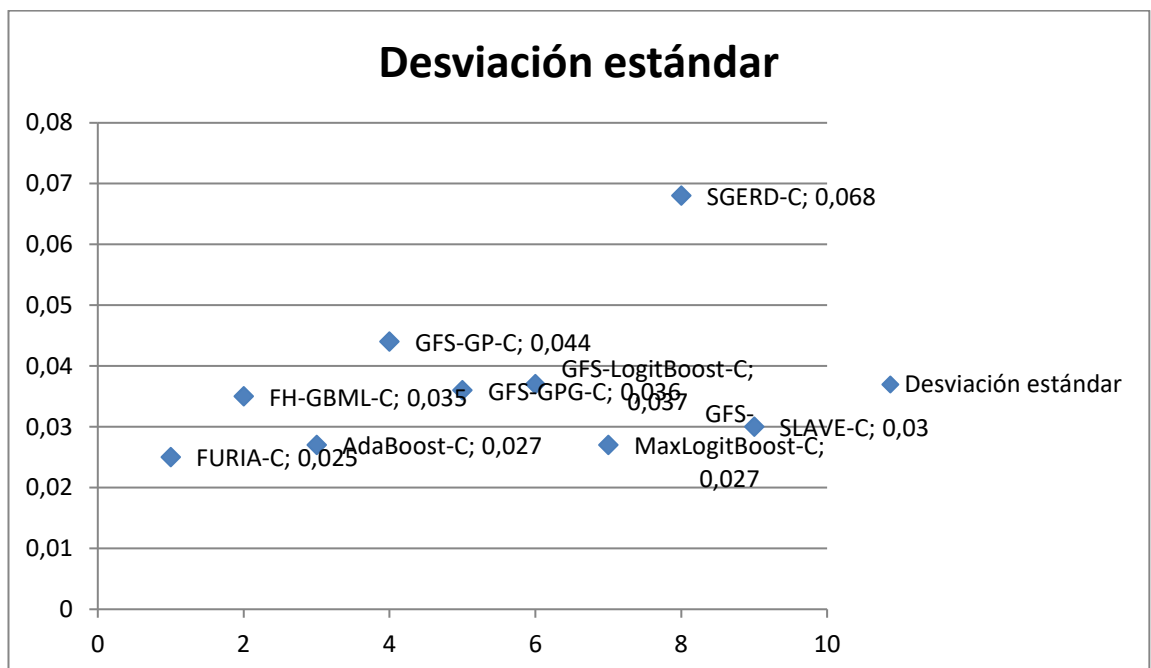


Ilustración 44.- Análisis de la desviación estándar

Se puede observar en la Ilustración 44 la desviación estándar de los 10 algoritmos de estudio detallados en la tabla 15 el algoritmo que posee mayor desviación estándar es el Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules From Data (SGERD-C) con un 0.068 frente al Furia siendo el algoritmo con menor desviación estándar con un 0.025

Tabla 15 Análisis de las desviaciones estándar

ITEM	ALGORITMO	%AC. TST
1	FURIA-C	0.025
2	FH-GBML-C	0.035
3	AdaBoost-C	0.027
4	GFS-GP-C	0.044
5	GFS-GPG-C	0.036
6	GFS-LogitBoost-C	0.037
7	GFS-MaxLogitBoost-C	0.027
8	SGERD-C	0.068
9	SLAVE-C	0.03

AUTORA: Olvera Toala Lely Johanna

FUENTE: investigación

4.1.3.2.3. Análisis de la precisión de los algoritmos difusos

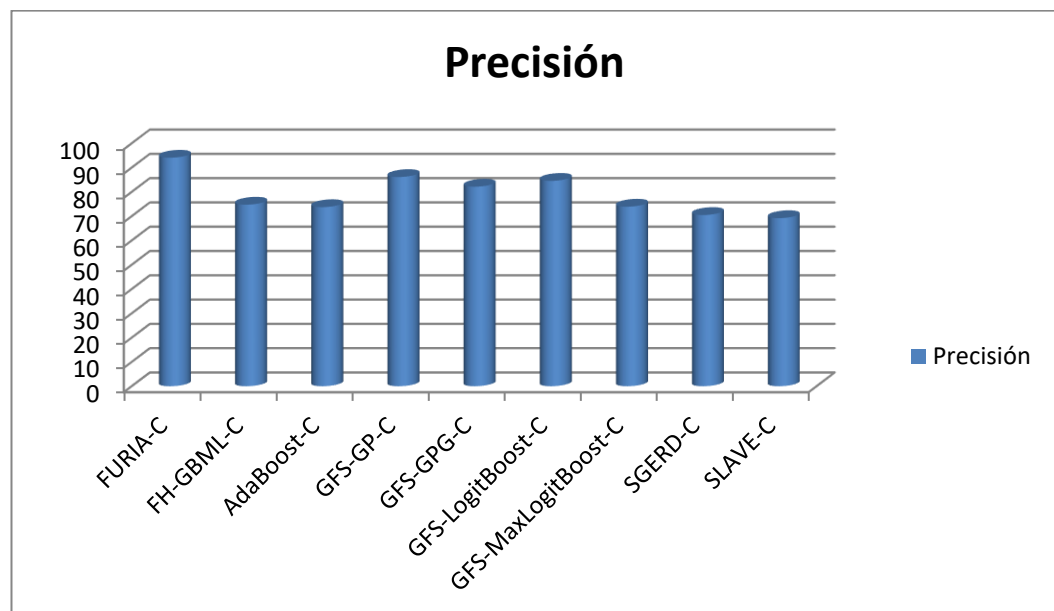


Ilustración 45.- Análisis de precisión

En la tabla 16 encontramos los algoritmos difusos aplicados en el trabajo de investigación, con apoyo en el gráfico 3 se puede analizar fácilmente que el algoritmo de mayor precisión es el FURIA y el de menor precisión es el AdaBoost por lo cual se concluye que como algoritmo que brinda mejor precisión es el algoritmo difuso FURIA.

Tabla 16 Análisis de datos de los Indices de precisión de los algoritmos difusos

ITEM	ALGORITMO	%AC. TST
1	FURIA-C	94,07
2	FH-GBML-C	74,74
3	AdaBoost-C	73,72
4	GFS-GP-C	86,12
5	GFS-GPG-C	82,14
6	GFS-LogitBoost-C	84.54
7	GFS-MaxLogitBoost-C	73,91
8	SGERD-C	70,48
9	SLAVE-C	69,19

AUTORA: Olvera Toala Lely Johanna

FUENTE: investigación

4.1.3.3. Análisis de la comprobación de la hipótesis mediante los algoritmos difusos

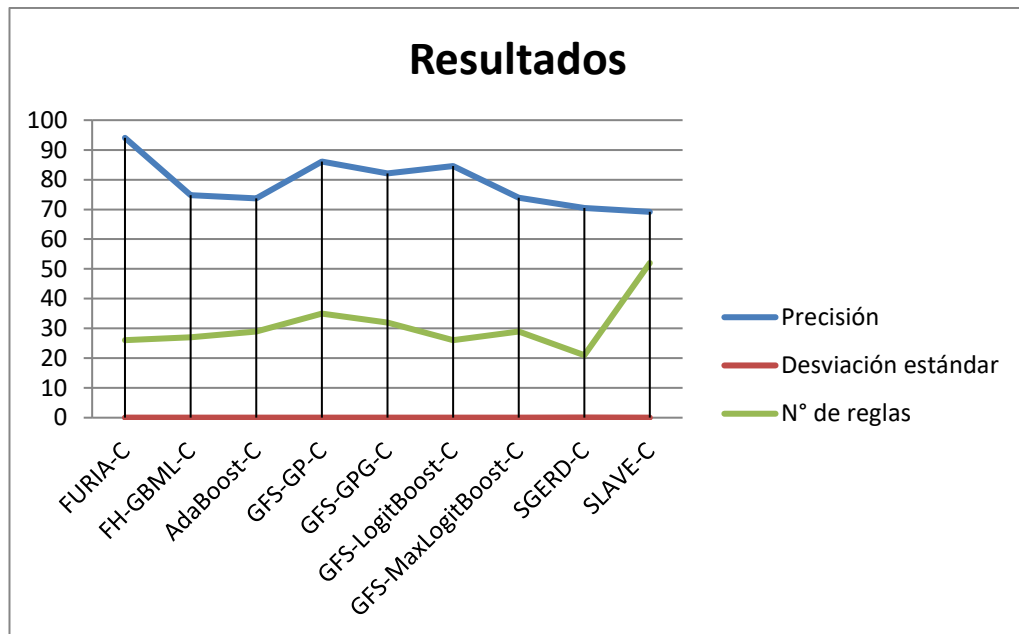


Ilustración 46.- Resultados

La Ilustración 46 de resultados se lo utiliza en la presente investigación como herramienta estadística para el análisis de los resultados obtenidos en el desarrollo de la misma en el cual se puede observar que de los 9 algoritmos aplicados, el algoritmo Steady-State Genetic Algorithm for Extracting Fuzzy Classification Rules From Data (SGERD-C) a pesar de contener una menor cantidad de reglas no es el mejor debido a que existen otros algoritmos con mayor precisión y menor desviación estándar.

El algoritmo difuso FURIA posee una mayor precisión frente a los otros algoritmos analizados lo que da la certeza de que posee mejores resultados, sin embargo se puede acotar que el algoritmo posee una desviación estándar menor que los otros algoritmos en comparación, lo que permite deducir que el algoritmo difuso FURIA es el algoritmo que se adapta mejor al caso de estudio, ya que las reglas son de una fácil interpretación como se muestra en la base de conocimiento anteriormente descrita.

A continuación se detalla un resumen de los resultados obtenidos en la investigación que tiene como tema: “Representación de conocimiento a través de sistemas difusos para evaluar los estilos de enseñanza/ aprendizaje entre estudiantes y docentes de la UTEQ”

Tabla 17.- Resultados de los algoritmos aplicados en Keel

ITEM	ALGORITMO	Nº REGLAS	%AC. TST (1)	%AC. TRN (2)
1	FURIA-C	26	94,07±0.025	96,39±0.006
2	FH-GBML-C	27	74,74±0.035	75,93±0.024
3	AdaBoost-C	29	73,72 ± 0.027	74,65± 0.023
4	GFS-GP-C	35	86,12±0.044	87,23±0.028
5	GFS-GPG-C	32	82,14±0.036	82,72±0.023
6	GFS-LogitBoost-C	26	84.54±0.037	88,65±0.022
7	GFS-MaxLogitBoost-C	29	73,91±0.027	73,76±0.017
8	SGERD-C	21	70,48±.068	71,17±0.046
9	SLAVE-C	52	69,19±0.003	69,19±0.304

AUTORA: Olvera Toala Lely Johanna

FUENTE: investigación

(1) Porcentaje de precisión y desviación estándar de pruebas

(2) Porcentaje de precisión y desviación estándar de entrenamiento

Se puede deducir según el análisis de la tabla 17 que el algoritmo FURIA brinda un porcentaje de precisión de 94,07% con 26 reglas buenas frente a los otros algoritmos que presentan valores de precisión más bajos y mayor número de reglas.

Siendo el análisis de resultados de la investigación con 26 reglas que dan lugar a la base de reglas que se detallan en el siguiente apartado de esta investigación.

4.1.3.4. Base de reglas

Después del desarrollo de la investigación que consiste en la extracción de conocimiento de los estilos de enseñanza/ aprendizaje entre docentes y discentes utilizando lógica difusa se construye la siguiente base de reglas que son el resultado del trabajo elaborado a través del proceso de KDD.

Tabla 18.- Base de reglas

N°	REGLA DIFUSA	PRECISIÓN
R1	Si le imparte clases más de 2 docentes con estilo de enseñanza auditivo, el estilo de aprendizaje es kinésico entonces Aprueba	93%
R2	Si el promedio de tercer año está entre 6.1 y 7.27, le imparten clases 2 docentes con estilos de enseñanza kinésico, el estilo de aprendizaje es visual y la estrategia de estudio es hacer cuadros comparativos entonces Aprueba	91%
R3	Si el promedio de tercer año está entre 6.1 y 7.27, el índice académico de primer año es inferior a 1.22, y al menos 1 docente de estilo de enseñanza visual le da clases entonces Aprueba	98%

R4	Si el promedio de tercer año está entre 7.27 y 8.12, le imparten clases 2 docentes con estilo de enseñanza kinésico, 1 con estilo de enseñanza visual y posee la estrategia de estudio Buscar y resumir ideas principales, y el estilo de aprendizaje es lectura entonces Aprueba	97%
R5	Si el Índice de aprobación de tercer año está entre 1.5 y 1.7, le dan clases al menos 2 docente con estilo de enseñanza kinésico, el índice de aprobación de primer año es inferior a 1.22, no posee hábitos de estudio y el estilo de aprendizaje es auditivo entonces Aprueba	96%
R6	Si estudia Gestión ambiental y el promedio de tercer año está entre 7.27 y 8.12, 1 docente con estilo de enseñanza visual y al menos le da 1 auditivo le imparten clases y su estilo de aprendizaje es kinésico entonces Aprueba	95%
R7	Si 2 docentes con el estilo de enseñanza kinésico le imparten clases, el Índice de aprobación de primer año está entre 1.02 y 1.75 y posee hábitos de estudio entonces Aprueba	99%

R8	Si el promedio de tercer año está entre 7.27 y 8.12 y el Índice académico de primer año es superior a 0.22 , ha tenido 2 docentes de estilo de enseñanza auditivo, 3 de kinésico y el estilo de aprendizaje es de tipo auditivo entonces Aprueba	99%
R9	Si el promedio de tercer año está entre 7.27 y 8.12 y le imparten 2 docentes con estilo de enseñanza kinésico y el Índice de aprobación de segundo año está entre 1.02 y 1.5 entonces Aprueba	95%
R10	Si el Índice de aprobación de tercer año es inferior 1.5, el índice académico de primer año es superior a 0.22 y le da clases al menos 1 docente con estilo de enseñanza visual y el estilo de aprendizaje es lectura entonces Aprueba	99%
R11	Si el promedio de tercer año está entre 7.27 y 8.12, le imparten clases 2 docentes con estilo de enseñanza kinésico y 3 auditivo, el promedio de segundo año está entre 7.14 y 7.95 y el índice de aprobación de tercero es inferior a 1.5 y su estilo de aprendizaje es auditivo entonces Aprueba	97%
R12	Si el promedio de tercer año es inferior a 7.27 y al menos 1 docente de estilo de enseñanza visual le da clases y aprende	95%

	de forma kinésica entonces Aprueba	
R13	Si el promedio de tercer año está entre 7.27 y 8.12, le imparten clases 2 docentes con estilos de enseñanza kinésico, 1 con visual y el índice de aprobación de primero es superior a 1.22 entonces Aprueba	95%
R14	Si el Índice de aprobación de tercer año está entre 1.5 y 1.7, le dan clases al menos 2 docente con estilo de enseñanza lectura, el índice de aprobación de primer año es superior a 1.22 entonces Aprueba	98%
R15	Si el índice de aprobación de primer año es inferior a 1.22, el índice académico de primer año está entre 0.25 y 0,50, le imparten al menos 2 docentes de estilo de enseñanza kinésico y 2 de visual, entonces Aprueba	96%
R16	Si el índice académico de primer año está entre 0.25 y 0.35, no le da ni un docente con estilos de enseñanza kinésico y auditivo pero si le imparten clases 2 docentes de tipo visual entonces Abandona sus estudios	96%

R17	Si el índice académico de primer año está entre 0.25 y 0.50, al menos 1 docente de estilo de enseñanza lectura y 2 de auditivo le imparten clases, y el tiempo de aprobación de segundo año es superior a 1.5 entonces Abandona sus estudios	97%
R18	Si la carrera es Marketing, le da clases1 docente con estilos de enseñanza lectura y el índice académico de primero es inferior 0.5 entonces Abandona sus estudios	93%
R19	Si el estilo de aprendizaje es kinésico y el tiempo de aprobación de segundo es inferior a 1.5, le imparten clases 2 docentes con el estilo de enseñanza auditivo y ni uno de kinésico entonces Abandona sus estudios	80%
R20	Si el promedio de tercer año es inferior a 7.27, no le imparten clases docente de estilo de enseñanza auditivo y el estilo de aprendizaje es auditivo entonces Abandona sus estudios	97%
R21	Si el estilo de aprendizaje es kinésico, le imparten clases 2 docentes de estilo de enseñanza auditivo y el promedio de tercer año es inferior a 7.27 entonces Abandona sus estudios	84%

R22	Si el índice académico de primer año es superior a 0.5 , el estilo de aprendizaje es auditivo entonces Aprueba	98%
R23	Si el promedio de tercer año es menor a 7.27, le imparten clases 1 docente con el estilo de enseñanza visual y su estilo de aprendizaje es Lectura entonces Aprueba	96%
R24	Si el promedio de tercer año es mayor a 7.27, el estilo de aprendizaje es auditivo, la estrategia de estudio es Resumir entonces Aprueba	88%
R25	Si le imparten clases 3 docentes con estilos de enseñanza visual, el aprendizaje es visual y el promedio de primer año es superior a 7.27 entonces aprobado	85%
R26	Si el promedio de tercer año es inferior a 7.27, el estilo de aprendizaje es lectura y le da 1 docente con estilo de enseñanza visual y 5 auditivos entonces Abandona sus estudios	84%

AUTORA: Olvera Toala Lely Johanna

FUENTE: investigación

4.2. DISCUSIÓN

Mediante el análisis de la tabla 15 se puede observar los diversos algoritmos difusos aplicados en la realización del presente estudio de investigación, es evidente que el algoritmo FURIA es el que brinda mejores resultados, basados en los porcentaje de precisión, desviación estándar y numero de reglas.

El algoritmo FURIA posee un índice de precisión de 94,07% con 26 reglas buenas frente a otros algoritmos que presentan valores de precisión más bajos y mayor número de reglas.

Permitiendo realizar un análisis de resultados de la base de conocimientos con 26 reglas que dan lugar a la etapa final del trabajo, en las que se puede evidenciar la relación que existe entre los estilos de enseñanza / aprendizaje, de manera especial entre el kinésico y el auditivo.

Según el trabajo de investigación” Sistemas de gestión de contenidos de enseñanza/ aprendizaje y técnicas de minería de datos para la enseñanza de ciencias computacionales: un caso de estudio en el norte de Coahuila” de (Olague, Torres, Morales, Valdez, & Silva, 2010) se puede observar que en los resultados obtenidos el estilo de enseñanza/ aprendizaje Kinésico es el que se adapta mejor al caso de estudio de ellos planteados teniendo similitud con el trabajo de investigación propuesto.

Por medio de la base de reglas descritas se puede evidenciar que la mayor parte de reglas poseen un alto índice de certeza lo que permite deducir que son reglas confiables, cabe recalcar que existen reglas que presentan los estilos de enseñanzas independiente de los de aprendizajes, y como toda investigación los resultados obtenidos son netamente lo que se refleja de los repositorios de datos con los que cuenta la Universidad Técnica Estatal de Quevedo.

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

Con la finalización de la investigación sobre la extracción de conocimiento a través de algoritmos difusos de los estilos de enseñanza / aprendizaje entre docentes y estudiantes de la Universidad Técnica Estatal de Quevedo se logró llegar a las siguientes conclusiones:

- Se logró establecer las variables que aportaban significativamente al modelo dando lugar a una vista minable de 13 atributos (**Tabla 9**), las mismas que permitieron la construcción de reglas que reflejan la relación entre los estilos de enseñanza/ aprendizaje.
- Mediante la aplicación de minería de datos por medio de técnicas de selección de atributos con el software Weka se logró analizar la consistencias de los datos, con la aportación de Keel se pudo discretizar los valores con el algoritmo “Uniform Frequency Discretizer”, e imputar datos con el algoritmo “Fuzzy K-means Imputation, permitiendo descubrir el conocimiento oculto de los datos con los que cuenta la Universidad Técnica Estatal de Quevedo, los cuales quedan a disponibilidad de la Unidad de planeamiento académico.
- Con la ejecución de algoritmos difusos por medio del software Keel se logró determinar que el algoritmo FURIA permitió obtener la base de conocimiento que establece las relaciones entre los estilos de enseñanza/ aprendizaje dando como resultado un total de 26 reglas (**Tabla 18**), debido al elevado índice de precisión y al conjunto de reglas de fácil de interpretación.

5.2. RECOMENDACIONES

Al finalizar el trabajo de investigación surgen las siguientes recomendaciones:

- Se recomienda seguir realizando trabajos investigativos con los datos almacenados en los gestores de base de datos de la Universidad Técnica Estatal de Quevedo para así poder descubrir conocimientos que aporten al desarrollo de la institución.
- Tener precaución con el ruido en los datos ya que podrían afectar al proceso de extracción de conocimiento, posiblemente ocasionando poca precisión de los algoritmos y reglas de difícil interpretación.
- Analizar la base de reglas obtenidas en la investigación por parte de la Unidad de Planeamiento Académico, y de ser posible ser consideradas las reglas que aporten a la toma de decisiones de la institución educativa.
- Extraer conocimiento oculto a partir de los repositorios de datos con los que cuenta la UTEQ en otros software tales como Analysis Service de SQL SERVER, KNIME, RapidMiner

CAPÍTULO VI: LITERATURA CITADA

6.1. BIBLIOGRAFÍA

- Alcalá, R; Alcala-Fdez, J; Casillas, J; Cordón, O; Herrera, F. (2006). Hybrid learning models to get the interpretabilityaccuracy trade-off in fuzzy modeling. *Soft Comput* 10(9), 717–734.
- Álvarez, M. (1994). *Fundamentos de Inteligencia Artificial*. Murcia: Universidad de Murcia.
- Arnold, & Blessie. (2013). Learning Styles of Teacher Education Students: Basis improving the Teaching. *Procedia - Social and Behavioral Sciences*. (págs. 3-11). Readcube.
- Ascolano, F., Cazorla, M., Alfonso, I., Colomina, O., & Lozano, M. (2003). *Inteligencia Artificial, modelos, técnicas y areas de aplicacion*. España: Thonsom.
- Ayyanathan, N., Kannammal, A., & Rekha, B. (May de 2012). Students' Communicative Competence Prediction and Performance Analysis of Probabilistic Neural Network Model. *IJCSI International Journal of Computer Science Issues*, Vol. 9, 302-317.
- Bernadó-Mansilla E; Ho TK. (2005). Domain of competence of XCS classifier system in complexity measurement space. *IEEE Trans Evol Comput* 9(1), 82–104.
- Bernal, C. A. (2010). *Metodología de la investigación*. Colombia: PEARSON EDUCACION.
- Bernal, L. (01 de Enero de 2010). *BernalPro*. Recuperado el 07 de 04 de 2015, de <http://bernal.pro/informatica/apuntesdeinformatica/finish/6-apuntes-de-informatica/83-mineria-de-datos>
- Brassard, Gilles; Bratle, Paul. (1997). *Fundamentos de Algoritmia*. Madrid : Prentice Hall.

- Cano JR; Herrera F; Lozano M. (2003). Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study. *IEEE Trans Evol Comput* 7(6), 561–575.
- Chellatamilan, T., Ravichandran, M., Suresh, R. M., & Kulanthaivel, G. (2011). Effect of Mining educational Data improve Adaptation of learning in e-Learning System. *Chennai and Dr.MGR University Second International Conference on Sustainable Energy and Intelligent System (SEISCON 2011)* , (págs. 922-927). India.
- Dais-ujat. (2007). *Avances en Informática y Sistemas Computacionales Tomo II*. Mexico: Universidad Autónoma de Tabasco.
- Duran, R., & Costaguta, R. (2007). Minería de datos para descubrir estilos de aprendizaje. *Iberoamericana*, 10.
- Dykinson. (2008). I Seminario sobre Sistemas Inteligentes. (pág. 256). Librería-Editorial Dykinson.
- Fayyad & Simoudis. (1997). Data mining and knowledge discovery. *Proceedings of international Conference on knowledge discovery and Data mining*, (págs. 3-16).
- Fleming, N., & D. Baume. (10 de 08 de 2011). *El cuestionario VARK. ¿Cómo aprendo mejor?* Recuperado el 04 de 21 de 2015, de : <http://www.vark-learn.com/documents/The%20VARK%20Questionnaire%20-%20Spanish.pdf>.
- Fraw-ley, P.-S. &. (1992). Knowledge Discovery in Databases: an overview. *Magazine*, 213-228.
- García, J. L.; Sánchez, C.; Jiménez, M. A.; Gutiérrez, M. (2012). Estilos de aprendizaje y estrategias de aprendizaje:. *Revista de Estilos deAprendizaje*, 65-78.
- Gilles Brassard;Paul Bratle. (1997). *Fundamentos de Algoritmos*. Madrid : Prentice Hall.

- González Peiteado, M. (2013). "Estilos de enseñanzas predominantes en los alumnos de los centros de formación del profesorado de la provincia de Pontevedra. *Revista Innovación Educativa*, 237-245.
- González-Peiteado, M. (2013). Los estilos de enseñanza y aprendizaje como soporte de la actividad docente. *Revista Estilos de aprendizaje*, 1-20.
- Goyal, M., & Vohra, R. (May de 2012). Applications of Data Mining in Higher Education. *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 2, No 1, 115-120.
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.
- Hawk, T. F., & Shah, A. J. (2007). Using learning style instruments to enhance student learning. *Decision Sciences of Innovative Education*, 1-19.
- J. Alcalá-Fdez; L. Sánchez; S. García; M.J. del Jesus; S. Ventura; J.M. Garrell; J. Otero; C. Romero; J. Bacardit; V.M. Rivas; J.C. Fernández; F. Herrera. (2009). KEEL: A Software Tool to Assess Evolutionary Algorithms to Data Mining Problems. *Soft Computing* 13, 307-318.
- Jimenez, M. N. (2008). *Como diagnosticar y mejorar los estilos de aprendizajes*. Santo Domingo: Asociacion Procompal.
- Knauf, R., Sakurai, Y., Takada, K., & Tsuruta, S. (2010). Personalizing learning processes by data mining. *Proceedings - 10th IEEE International Conference on Advanced Learning Technologies*, (págs. 488-492).
- Laal, K.-K., & Laal, M. (2014). Teaching and education; collaborative style. *5th World Conference on Educational Sciences - WCES 2013*, Procedia - Social and Behavioral Sciences 116 4057- 4061.
- Lozano R., A. (2001). *Estilos de aprendizaje y enseñanza*. México: Trillas.
- MacGregor, J. (1990). "Collaborative Learning: Shared Inquiry as a Process of Reform.". *New Directions for Teaching and Learning* no. 42, 19-30.

- Martín, J. D. (2000). *Implementación de Redes Neuro-Difusas Para Per Aplicadas en Problemas de Clasificación y Modelación*. USA: Dissertation.
- Martínez, F. J., & López, J. C. (2010). *Sistemas inteligentes de marketing para modelado causal*. Granada, España: DELTA.
- Martínez-Estudillo A; Martínez-Estudillo F; Hervás-Martínez C; García-Pedrajas N. (2006). Evolutionary product unit based neuralnetworks for regression. *Neural Netw* 19, 477–486.
- Orientacion Andujar. (29 de 05 de 2014). *Orientacionandujar*. Recuperado el 21 de 04 de 2015, de <http://www.orientacionandujar.es/2014/05/29/test-de-estilos-de-aprendizaje-de-test-de-vark/>
- Osmanbegovic, E., Agic, H., & Suljic, M. (2014). Prediction of Students'Success by Applying. *Journal of Theoretical and Applied Information Technology*, 381-387.
- Parack, Suhem; Zahid, Zain; Merchant,Fatima. (2012). Application of data mining in educational databases for predicting academic trends and patterns. *In Technology Enhanced Education (ICTEE), 2012 IEEE International Conference*, (págs. 1-4).
- Priya , K. S., & Kumar, A. (2013). Improving the Student's Performance Using Educational Data Mining. *Int. J. Advanced Networking and Applications* , 1680-1685.
- Puente, P. (1996). *La teoria de subconjuntos borrosos: Aplicaciones*. Asturias: Universidad de Oviedo.
- Rob, P., & Coronel, C. (2004). Minería de Datos. En R. Peter, & C. Coronel, *Sistemas de bases de datos: Diseño, implementación y administración* (pág. 654). Mexico.
- Sicilia, C. Á., & Delgado, N. M. (2002). Educación Física y Estilos de Enseñanza. *Publicaciones: INDE. Barcelona-España*.

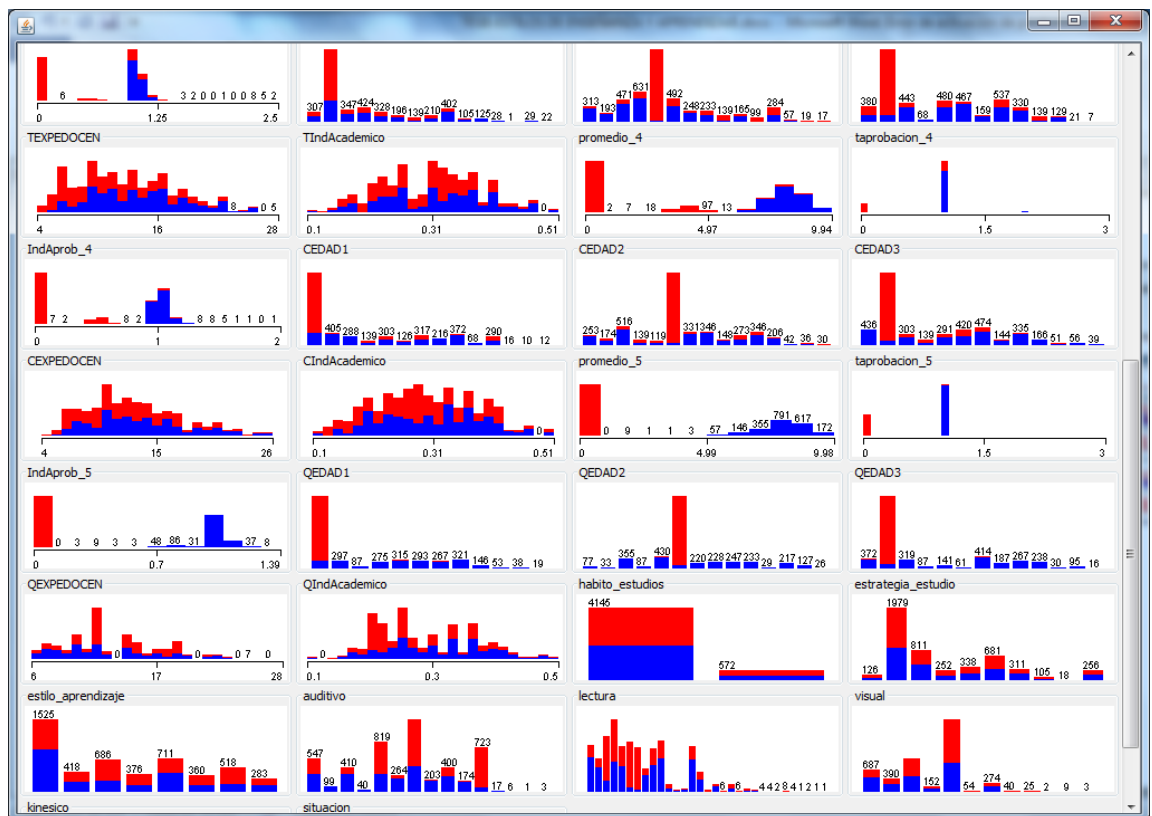
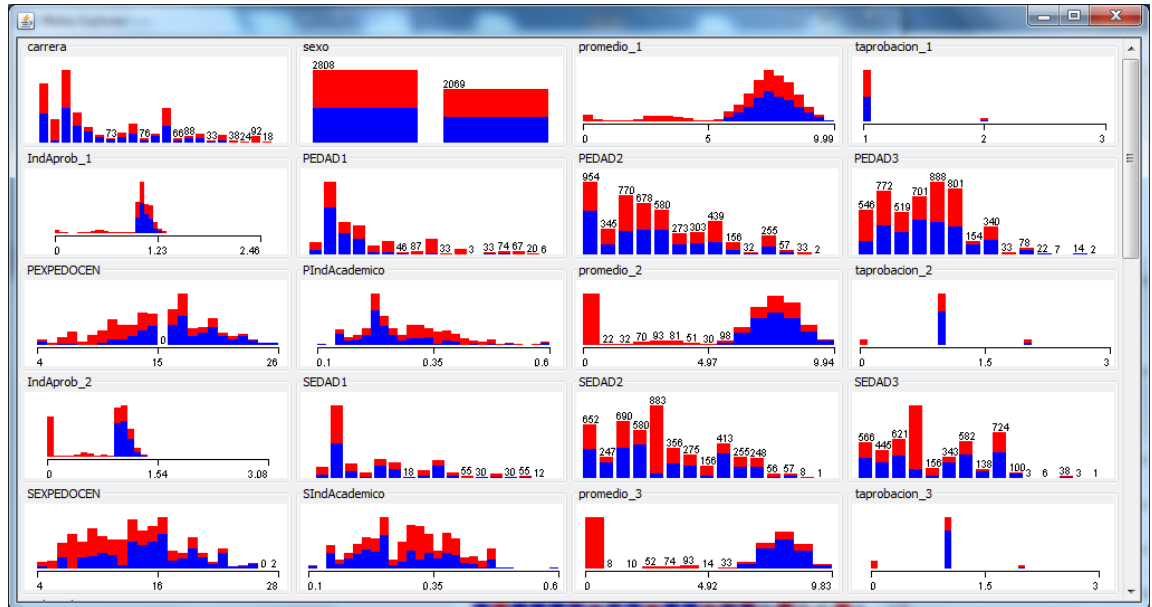
- Singh, Chandrani; Gopal, Arpita; Mishra, Santosh. (2011). Extraction and analysis of faculty performance of management discipline from student feedback using clustering and association rule mining techniques. *In Electronics Computer Technology (ICECT), 2011 3rd International Conference*, (págs. 94-96).
- Siraj, Fadzilah; Abdoulha, Mansour Ali. (2009). Uncovering hidden information within university's student enrollment data using data mining. *Third Asia International Conference*, (págs. 413-418).
- Soler, V. (2007). *Lógica difusa aplicada a conjuntos imbalanceados*.
- Vialardi, C., Chue, J., Peche, J. P., Alvarado, G., Vinatea, B., Estrella, J., y otros. (2011). A data mining approach to guide students through the enrollment process based on academic performance. *Springer Science+Business Media*(21), 217-248.
- Xiaolong Zhang; Guirong Liu. (2008). Score Data Analysis for Pre-Warning Students in University. *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference* , (págs. 1-4).
- Zhang, J. M., & Gao, W. X. (2008). Application of association rules mining in the system of university teaching appraisal. *International Workshop on Education Technology and Training and 2008 International Workshop on Geoscience and Remote Sensing, ETT and GRS*, (págs. 26-28).
- Zhang, Z. (2010). Study and analysis of data mining technology in college courses students failed. *In 2010 International Conference on Intelligent Computing and Integrated Systems*, (págs. 800-802).
- Zhi-min, Y., Qing, S., & Bin, S. (2010). The Analysis of Student's Grade Based on Rough Sets. *3rd IEEE International Conference on Ubi-Media Computing*, (págs. 345-349).

Zwilling, M., & Natek, S. (2013). Data Mining for small Student Data Set-Knowledge Management System for Higher Education Teachers. *Expert Systems with Applications*, Vol. 41, 1380-1388.

CAPÍTULO VII: ANEXOS

7.1. ANEXOS

Vista de todos los datos trabajados en weka



CUESTIONARIO VARK



Va a elegir sus alimentos en un restaurante o café. Ud.:

- ☐ Elegiría algo que ya ha probado en ese lugar.
- ☐ Escucharía al mesero o pediría recomendaciones a sus amigos.
- ☐ Observaría lo que otros están comiendo o las fotografías de cada platillo.
- ☐ Elegiría a partir de las descripciones del menú.

Un sitio web tiene un video que muestra cómo hacer un gráfico especial. Hay una persona que habla, algunas listas y palabras que describen lo que debe hacer y algunos diagramas. Se podría aprender más de:

- ☐ Ver los diagramas.
- ☐ La lectura de las palabras.
- ☐ Viendo las acciones.
- ☐ Escuchando.

Va a cocinar algún platillo especial para su familia. Ud.:

- ☐ Pediría sugerencias a sus amigos.
- ☐ Hojearía un libro de cocina para tomar ideas de las fotografías.
- ☐ Cocinaría algo que conoce sin la necesidad de instrucciones.
- ☐ Utilizaría un libro de cocina donde sabe que hay una buena receta.

Le gustan los sitios web que tienen:

- ☐ Canales de audio para oír música, programas o entrevistas.
- ☐ Descripciones escritas interesantes, características y explicaciones.
- ☐ Cosas que se pueden picar, mover o probar.
- ☐ Un diseño interesante y características visuales.

Tiene un problema con su rodilla. Preferiría que el doctor:

- ☐ Le diera una dirección web o algo para leer sobre el asunto.
- ☐ Le describiera qué está mal.
- ☐ Utilizará el modelo plástico de una rodilla para mostrarle qué está mal.

- ☐ Le mostrara con un diagrama qué es lo que está mal.

Está utilizando un libro, CD o sitio web para aprender cómo tomar fotografías con su nueva cámara digital. Le gustaría tener:

- ☐ Instrucciones escritas con claridad, con características y puntos sobre qué hacer.
- ☐ Muchos ejemplos de fotografías buenas y malas y cómo mejorar éstas.
- ☐ Diagramas que muestren la cámara y qué hace cada una de sus partes.
- ☐ La oportunidad de hacer preguntas y que le hablen sobre la cámara y sus características.

Además del precio, ¿qué influiría más en su decisión de comprar un nuevo libro de no ficción?

- ☐ La apariencia le resulta atractiva.
- ☐ Un amigo le habla del libro y se lo recomienda.
- ☐ Tiene historias, experiencias y ejemplos de la vida real.
- ☐ Una lectura rápida de algunas partes del libro.

Prefiere a un profesor o un expositor que utiliza:

- ☐ Demostraciones, modelos o sesiones prácticas.
- ☐ Diagramas, esquemas o gráficas.
- ☐ Folletos, libros o lecturas.
- ☐ Preguntas y respuestas, charlas, grupos de discusión u oradores invitados.

Está a punto de comprar una cámara digital o un teléfono móvil. ¿Además del precio, qué más influye en su decisión?

- ☐ Los comentarios del vendedor acerca de las características del aparato.
- ☐ El diseño del aparato es moderno y parece bueno.
- ☐ La lectura de los detalles acerca de las características del aparato.
- ☐ Lo utiliza o lo prueba.

Recuerde la vez cuando aprendió cómo hacer algo nuevo. Evite elegir una destreza física, como montar bicicleta. ¿Cómo aprendió mejor?:

- ☐ Siguiendo instrucciones escritas en un manual o libro de texto.

- ☐ Siguiendo pistas visuales en diagramas y gráficas.
- ☐ Escuchando la explicación de alguien y haciendo preguntas.
- ☐ Viendo una demostración.

Está ayudando a una persona que desea ir al aeropuerto, al centro de la ciudad o a la estación del ferrocarril. Ud.:

- ☐ Iría con ella.
- ☐ Le diría cómo llegar.
- ☐ Le daría las indicaciones por escrito (sin un mapa).
- ☐ Le daría un mapa.

Un grupo de turistas desea aprender sobre los parques o las reservas de vida salvaje en su área. Ud.:

- ☐ Les daría libros o folletos sobre parques o reservas de vida salvaje.
- ☐ Les mostraría figuras de Internet, fotografías o libros con imágenes.
- ☐ Los llevaría a un parque o reserva y daría una caminata con ellos.
- ☐ Les daría una plática acerca de parques o reservas de vida salvaje.

Ha acabado una competencia o una prueba y quisiera una retroalimentación. Quisiera tener la retroalimentación:

- ☐ Escuchando a alguien haciendo una revisión detallada de su desempeño.
- ☐ Utilizando gráficas que muestren lo que ha conseguido.
- ☐ Utilizando ejemplos de lo que ha hecho.
- ☐ Utilizando una descripción escrita de sus resultados.

Tiene que hacer un discurso importante para una conferencia o una ocasión especial. Ud.:

- ☐ Escribiría su discurso y se lo aprendería leyéndolo varias veces.
- ☐ Elaboraría diagramas o conseguiría gráficos que le ayuden a explicar las ideas.
- ☐ Conseguiría muchos ejemplos e historias para hacer la charla real y práctica.
- ☐ Escribiría algunas palabras clave y practica su discurso repetidamente.

Desea aprender un nuevo programa, habilidad o juego de computadora. Ud. debe:

- ☐ Seguir los diagramas del libro que vienen con el programa.
- ☐ Platicar con personas que conocen el programa.
- ☐ Utilizar los controles o el teclado.
- ☐ Leer las instrucciones escritas que vienen con el programa.

Está planeando unas vacaciones para un grupo de personas y desearía la retroalimentación de ellos sobre el plan. Ud.:

- ☐ Describiría algunos de los atractivos del viaje.
- ☐ Les llamaría por teléfono, les escribiría o les enviaría un e-mail.
- ☐ Utilizaría un mapa o un sitio web para mostrar los lugares.
- ☐ Les daría una copia del itinerario impreso.