



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA

Carrera
INGENIERÍA EN SISTEMAS

Tema de la Tesis

“Estudio de caracterización de alumnos graduados de la UTEQ mediante minería de datos educacional”

Previo a la obtención del título de:

Ingeniero en Sistemas

Autor:

Iván Darío Ponce Vélez

Director de Tesis

PhD. Amilkar Puris Cáceres

Quevedo – Ecuador

2015



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA

Carrera
INGENIERÍA EN SISTEMAS

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

Aprobado:

PRESIDENTE DEL TRIBUNAL DE TESIS

Ing. Msc. Carlos Márquez

MIEMBRO DEL TRIBUNAL DE TESIS

Ing. Msc. Washington Chiriboga

MIEMBRO DEL TRIBUNAL DE TESIS

Ing. Msc. José Luís Tubay Vergara

DEDICATORIA

El presente trabajo lo dedico a mi Padres, María Natividad y Rubén Darío, a quien le agradezco por su paciencia y comprensión, prefirieron sacrificar su tiempo para que yo pudiera cumplir con el mío. Por su bondad y sacrificio me inspiraron ser mejor cada día, ahora puedo decir que esta tesis lleva mucho de ellos, gracias por estar a mi lado.

A mis hermanos, María Elena y Luís Alfredo, por brindarme su apoyo incondicional, motivarme y ayudarme a cumplir este objetivo.

A mi pequeño sobrino Ethan Sebastián, quien con su sonrisa y espontaneidad llena de alegría la vida de quienes lo rodemos.

Iván Darío Ponce Vélez

AGRADECIMIENTOS

En primer lugar agradezco a DIOS, por permitirme cumplir una de mis principales metas y por bendecirme con las personas que tengo a mi alrededor, quienes son parte importante en cada etapa de mi vida.

Agradezco a mis amigos por estar a mi lado en los momentos más difíciles de mi corta vida y brindarme su valiosa compañía durante este caminar.

Agradezco de manera especial y personal al Ing. Jorge Guanín y al PhD. Amilkar Puris por brindar desinteresadamente su conocimiento como aporte fundamental para lograr culminar el presente trabajo de Investigación.

Y a todas y cada una de las personas que con sus diferentes aportes me han ayudado a que culmine con éxito el presente Proyecto de Investigación.

Iván Darío Ponce Vélez

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, IVÁN DARÍO PONCE VÉLEZ, 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.

Iván Darío Ponce Vélez

CERTIFICACIÓN DEL DIRECTOR DE TESIS

El suscrito, Amilkar Puris Cáceres, Docente de la Universidad Técnica Estatal de Quevedo, certifica que el Egresado IVÁN DARÍO PONCE VÉLEZ, realizó la tesis de grado previo a la obtención del título de INGENIERO EN SISTEMAS titulada **“ESTUDIO DE CARACTERIZACIÓN DE ALUMNOS GRADUADOS DE LA UTEQ MEDIANTE MINERÍA DE DATOS EDUCACIONAL”** bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

Phd. Msc. Amilkar Puris Cáceres

ÍNDICE

| | |
|---------------------------------------------------------------------------------------|--------------------------------------|
| ÍNDICE DE TABLAS | x |
| ÍNDICE DE ILUSTRACIONES | xi |
| RESUMEN EJECUTIVO | xii |
| ABSTRACT..... | xiii |
| CAPÍTULO I - MARCO CONTEXTUAL DE LA INVESTIGACIÓN | 1 |
| 1.1. INTRODUCCIÓN | 2 |
| 1.2. JUSTIFICACIÓN | 4 |
| 1.3. PROBLEMATIZACIÓN | 5 |
| 1.3.1. Análisis del problema | 5 |
| 1.3.2. Formulación | 5 |
| 1.3.3. Sistematización del Problema | 6 |
| 1.4. OBJETIVOS | 7 |
| 1.4.1. Objetivo General | 7 |
| 1.4.2. Objetivos Específicos..... | 7 |
| 1.5. HIPÓTESIS | 8 |
| 1.5.1. Planteamiento | 8 |
| 1.5.2. Matriz de Operacionalización | 8 |
| CAPÍTULO II - MARCO TEÓRICO..... | 9 |
| 2.1. FUNDAMENTACIÓN CONCEPTUAL | 10 |
| 2.1.1. KDD (Knowledge Discovery in Database)..... | 10 |
| 2.1.2. Minería de datos | 11 |
| 2.1.3. Minería de datos educacional..... | 11 |
| 2.1.4. Soft Computing | 11 |
| 2.1.5. CRISP-DM | 12 |
| 2.1.6. Técnicas predictivas..... | ¡Error! Marcador no definido. |
| 2.1.7. Técnicas descriptivas..... | 12 |
| 2.1.7.1. Técnicas algebraicas y estadísticas | 12 |
| 2.1.7.2. Técnicas bayesianas | 13 |
| 2.1.7.3. Técnicas basadas en conteos de frecuencias y tablas de contingencia | 13 |

| | | |
|------------------------------------------------------|-----------------------------------------------------------------------------------|----|
| 2.1.7.4. | Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas | 14 |
| 2.1.7.5. | Técnicas relacionales y estructurales | 14 |
| 2.1.7.6. | Técnicas basadas en redes neuronales artificiales | 14 |
| 2.1.7.7. | Técnicas basadas en núcleo y máquinas de soporte vectorial..... | 15 |
| 2.1.7.8. | Técnicas estocásticas y difusas | 15 |
| 2.1.8. | Software Keel | 16 |
| 2.1.9. | Software Knime | 17 |
| 2.1.10. | Software Rapid Miner | 17 |
| 2.2. | MARCO REFERENCIAL | 19 |
| CAPÍTULO III - METODOLOGÍA DE LA INVESTIGACIÓN | | 21 |
| 2.3. | MATERIALES Y MÉTODOS | 22 |
| 2.3.1. | Equipos y Materiales..... | 22 |
| 2.3.2. | Métodos y Técnicas utilizados en la Investigación | 23 |
| 2.3.2.1. | Tipo de Investigación..... | 24 |
| 2.3.2.2. | Técnicas de Investigación..... | 25 |
| 2.3.3. | Población y Muestra | 26 |
| CAPÍTULO IV - RESULTADOS Y DISCUSIÓN | | 27 |
| 4.1. | RESULTADOS | 28 |
| 4.1.1. | Preparación de la función para extraer los datos..... | 28 |
| 4.1.2. | Preparación de los datos. | 29 |
| 4.1.2.1. | Transformación y selección de atributos | 33 |
| 4.1.2.2. | Transformación de atributos | 39 |
| 4.1.2.3. | Atributos seleccionado de las diferentes vistas minables | 40 |
| 4.1.2.4. | Tratamiento de valores faltantes | 41 |
| 4.1.2.4.1. | Imputación de datos | 42 |
| 4.1.2.5. | Técnicas de discretización..... | 43 |
| 4.1.3. | Extracción de conocimiento | 43 |
| 4.1.3.1. | Resultados | 44 |
| 4.1.3.2. | Base de reglas. | 46 |
| 4.1.3.3. | Análisis de las Resultados | 48 |
| 4.1.3.3.1. | Interpretación del gráfico | 48 |

| | |
|------------------------------------------------------------------------------------|----|
| 4.1.3.3.1.1. Interpretación del gráfico con respecto a la precisión... | 48 |
| 4.1.3.3.1.2. Interpretación del gráfico con respecto a la Desviación estándar..... | 49 |
| 4.1.3.3.1.3. Interpretación del gráfico con respecto al Número de reglas..... | 50 |
| 4.2. DISCUSIÓN..... | 51 |
| CAPÍTULO V - CONCLUSIONES Y RECOMENDACIONES | 53 |
| 5.1. CONCLUSIONES | 54 |
| 5.2. RECOMENDACIONES | 56 |
| CAPÍTULO VI - BIBLIOGRAFÍA..... | 57 |
| 6.1. LITERATURA CITADA | 58 |
| CAPÍTULO VII - ANEXOS..... | 62 |

ÍNDICE DE TABLAS

| | |
|------------------------------------------------------------------|----|
| Tabla 1. Matriz de Operacionalización..... | 8 |
| Tabla 2. Equipos y Materiales | 22 |
| Tabla 3. Matriz de Atributos | 33 |
| Tabla 4. Atributos Seleccionados - Vista Minables | 41 |
| Tabla 5. Algoritmos Implementados | 45 |
| Tabla 6. Resultados aplicando los algoritmos implementados. | 45 |
| Tabla 7. Bases de reglas..... | 47 |

ÍNDICE DE ILUSTRACIONES

| | |
|-----------------------------------------------------------------|----|
| Ilustración 1. Herramienta Weka | 34 |
| Ilustración 2. Módulo de Selección de Atributos - Weka | 35 |
| Ilustración 3. Diagrama de selección de atributos. Knime | 36 |
| Ilustración 4. Atributos otorgado por el Software | 37 |
| Ilustración 5. Esquema genérico de algoritmo tipo wrapper | 38 |
| Ilustración 6. Esquema genérico de algoritmo tipo filtro | 38 |
| Ilustración 7. Pseudocódigo de un algoritmo híbrido | 38 |
| Ilustración 8. Algoritmos Implementados en Keel | 43 |
| Ilustración 9. Resultados | 48 |

RESUMEN EJECUTIVO

El presente trabajo investigativo busca encontrar el conocimiento oculto a través de minería de datos educacional para obtener una caracterización de graduados en la Universidad Técnica Estatal de Quevedo.

Mediante la gran cantidad de información alojada en los repositorios de datos de la Universidad y el análisis empírico por parte de las autoridades surge la necesidad de realizar esta investigación, con el fin de encontrar el conocimiento oculto entre una montaña de información.

El presente trabajo tiene como objetivo descubrir los factores que inciden en los perfiles de estudiantes graduados, analizando los datos desde el periodo inicial de estudio hasta su culminación. Mediante minería de datos educacional.

Los resultados de esta investigación se pretenden otorgar a la Unidad de Planeamiento Académico para que en base a su experiencia puedan realizar un mejor análisis de la información, además de que sirva como soporte idóneo en el proceso de toma de decisiones de la unidad antes mencionada.

El autor de la investigación detalla el proceso o desarrollo del presente trabajo en siete capítulos, siendo el capítulo IV donde presenta detalladamente la manera técnica de cómo se obtuvo el conocimiento, partiendo de la ingente cantidad de información que otorgó la institución.

Los resultados entregados, un conjunto de reglas fueron seleccionados en base a los niveles de confianza, desviación estándar y número de reglas otorgadas por parte de cada uno de los algoritmos aplicados en este caso de estudio.

ABSTRACT

This research work aims to find the hidden knowledge through educational data mining to obtain a characterization of graduates from Universidad Técnica Estatal de Quevedo.

Through the many housed in data repositories college and empirical analysis by the authorities with information arises the need for this research, in order to find the hidden knowledge from a mountain of information.

This paper aims to describe the factors that influence the profiles of graduates, analyzing data from the initial study period to completion. Means educational data mining.

The results of this research aims to give the Unit for Academic Planning based on experience can make a better analysis of information, in addition to serving as suitable support in the decision making process of the aforementioned unit.

The author of the research detailed the process or application of this work in seven chapters, Chapter IV which presents in detail the technical way of how knowledge is obtained starting from the wealth of information that granted the institution.

The results provided a set of rules were selected based on the confidence levels, standard deviation and number of rules issued by pate of each of the algorithms applied in this case study.

CAPÍTULO I

MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1. INTRODUCCIÓN

Información es un conjunto organizado de datos debidamente procesado que constituye un mensaje real sobre un determinado hecho o fenómeno, por esta razón se dice que “la información es poder”, sin embargo existen grandes almacenes de datos que carecen de un estudio adecuado para poder convertir ese conjunto de datos en información útil y aún más, en conocimiento.

En la actualidad se emplea el proceso de minería de datos para un mejor y completo análisis de datos con la finalidad de descubrir características, relaciones, dependencias o tendencias de los datos previamente desconocidas (Rob & Coronel, 2004). El Data Mining permite encontrar patrones escondidos que son de gran utilidad para plantear estrategias que apunten a los objetivos del negocios (Canney, 2007), siendo implementada en numerosos ambientes de desarrollo, entre estos la educación.

Al introducir este término de minerías de datos en la educación, surge un nuevo término: Minería de datos educacional EDM (siglas en inglés) es una disciplina emergente, preocupada por el desarrollo de métodos para explorar los tipos únicos de los datos que provienen de los centros educativos, y el uso de esos métodos para entender mejor a los estudiantes (International Educational Data Mining Society, 2011), para lograr una significativa mejora en los procesos de toma de decisiones en el ámbito educativo.

EDM enfoca sus procesos en la manera de aplicar la minería de datos para el análisis de los datos generados por los diferentes sistemas de información que apoyan el aprendizaje en la educación.

Entonces, una manera de afrontar estas necesidades que tiene una institución de educación superior es mediante la minería de datos educacional. Esta permite a las entidades educativas descubrir y conocer patrones ocultos en sus grandes almacenes de datos. A continuación, estos patrones se incorporan en modelos de minería de datos y se utilizan para asociar las conductas

individuales con gran precisión. Por ejemplo, la minería de datos puede ofrecer a una institución de tercer nivel la información necesaria para tomar medidas antes de que un estudiante abandone o para asignar de manera eficaz los recursos a una estimación precisa del número de estudiantes que se matricularán en un determinado curso. Como resultado de este tipo de información, dichas instituciones pueden asignar los recursos y su personal de manera más eficaz.

Sin embargo, en la actualidad se desconoce que se haya aplicado un estudio de caracterización de estudiantes de la UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO (UTEQ) con ayuda de minería de datos educacional dando como resultado un conocimiento oculto que surge como consecuencia de haber aplicado razonamiento aproximado que es el conjunto de técnicas y metodologías para el tratamiento de información imprecisa o incierta (VALVERDE, 1992).

La UTEQ maneja grandes cantidades de datos, no obstante hace falta un adecuado tratamiento de este gran volumen de datos, con el fin de entregar información válida y confiable que permita a la Unidad de Planeamiento Académico (UPA) una mejora en la toma de decisiones para un excelente desarrollo académico en la institución.

Un adecuado análisis inteligente de datos, el cual es una área de la informática que busca generar o adquirir conocimiento (PRIESCA, s.f.) permitirá encontrar el conocimiento oculto para obtener un perfil de estudiantes graduados y poder predecir o asociar las tendencias de estudio de los alumnos y entregar esa información para que pueda ser analizada por la UPA logrando contrarrestar el bajo índice de estudiantes graduados que oferta la UTEQ según informes entregados por la secretaría de las facultades internas de la institución.

1.2. JUSTIFICACIÓN

Con el propósito de realizar un adecuado estudio de los factores que intervienen en la culminación exitosa de los estudios de un alumno de pregrado en la UTEQ, para ofrecer un soporte eficaz que permitan una asistencia apropiada en la gestión académica y una mejora continua en la calidad de estudio que brinda esta universidad como entidad de educación superior.

El hecho de que un estudiante puede culminar sus estudios de pregrado o no, además de saber qué tiempo le toma graduarse, no solo depende del rendimiento del mismo, pueden intervenir muchos otros factores en este fenómeno, lo que da lugar para realizar un estudio que nos permita identificar este tipo de estudiante a través de uno o varios perfiles.

La obtención de un perfil de estudiantes graduados a través de técnicas supervisadas y no supervisadas permitirá mejorar la gestión académica de la UTEQ, al poder predecir el éxito de los estudios universitarios de un determinado estudiante o bien, asociar las diferentes características de un alumno con el fin de mejorar la toma de decisiones y mejorar el índice de graduados de la institución.

Esta investigación tiene lugar en la UTEQ en el departamento de las TIC's (Tecnologías de la Información y la Comunicación) del Instituto de Informática donde se encuentra alojado los SGDB que entregaran el conjunto de datos de los cuales se realizará el análisis de los estudiantes que se matricularon entre los periodos del año 2003 y el año 2008.

1.3. PROBLEMATIZACIÓN

1.3.1. Análisis del problema

En un mundo caracterizado por rápidos y complejos cambios tecnológicos, a los directivos universitarios cada vez se les plantea problemas más difíciles que no solo afectan a su institución educativa, sino que influyen además en el conjunto de la sociedad.

Por esto, cada vez tendrán una mayor necesidad de enfrentar las situaciones con ayuda de nuevas tecnologías. Sin embargo, en la U.T.E.Q se dificulta la toma de decisiones por parte de la Unidad de Planeamiento Académico (UPA) debido a que carecen de un buen estudio de los datos que se encuentran en los SGBD (Sistemas Gestores de Bases de Datos).

Según informes entregados por las secretarías de las facultades de la institución educativa, menos de un quintil de la población estudiantil logra culminar con éxitos su carrera profesional, generando un gran problema en el indicador resultante de la relación de los estudiantes matriculados a las carreras universitarias entre los que lograron graduarse.

Analizar este gran volumen de datos que nos entrega el SGDB de la U.T.E.Q de una manera empírica resulta extremadamente tedioso para el personal administrativo de la institución, además de ofrecer un análisis poco acertado a la realidad y sin bases que fomenten su propuesta.

1.3.2. Formulación

¿Es aplicable la minería de datos educacional para determinar los factores que influyen en la graduación de un estudiante de la UTEQ?

1.3.3. Sistematización del Problema

- ¿Cuáles son los atributos que logren proporcionar información significativa sobre el desempeño estudiantil?
- ¿Qué técnicas de minería de datos permiten obtener un significativo aporte positivo a los resultados finales de la investigación?
- ¿Se obtendrá un conjunto de reglas que permita proporcionar un escenario idóneo para identificar a un estudiante graduado?

1.4. OBJETIVOS

1.4.1. Objetivo General

Aplicar minería de datos educacional para determinar los factores que inciden en los perfiles de estudiantes graduados de la UTEQ.

1.4.2. Objetivos Específicos

- Identificar los atributos que proporcionen información significativa sobre el desempeño estudiantil.
- Determinar las técnicas de minería de datos que permitan obtener un aporte significativamente positivo en los resultados, el cual nos ayude en el proceso de toma de decisiones de la unidad de planeamiento académico de la UTEQ.
- Definir un conjunto de reglas que permitan proporcionar el mejor escenario para identificar a un estudiante graduado.

1.5. HIPÓTESIS

1.5.1. Planteamiento

El análisis de muchos factores de carácter exógeno y endógeno permitirá obtener un perfil de éxito de los estudios de pregrado de un alumno de la UTEQ.

1.5.2. Matriz de Operacionalización

| Tipos de variable | Definición conceptual | Dimensión de la variable | Indicador |
|----------------------------------------------|--------------------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| Independiente | | | |
| Factores exógenos y endógenos del estudiante | Características internas y externas que pertenecen a un estudiante | <ul style="list-style-type: none">• Modelación | <ul style="list-style-type: none">• Atributos• Técnicas• Reglas |
| Dependiente | | | |
| Perfil de éxito de estudiantes graduados | Conjunto de características que cumple un determinado estudiante | <ul style="list-style-type: none">• Confiabilidad | <ul style="list-style-type: none">• Porcentaje de precisión• Índice de Cobertura |

Tabla 1. Matriz de Operacionalización

CAPITULO II

MARCO TEÓRICO

2.1. FUNDAMENTACIÓN CONCEPTUAL

2.1.1. KDD (Knowledge Discovery in Database)

Este proceso consiste en la extracción de conocimiento partiendo de fuente masivas de datos que además se encarga de la preparación de los datos y de los resultados obtenidos. KDD se ha definido como el proceso no trivial de identificación de los datos de patrones válidos de las bases de datos para interpretar grandes cantidades de información y encontrar las relaciones o patrones entre los datos, que se encuentran en la base de datos (Molina, 2007).

El KDD se define como un proceso no trivial de extracción de patrones válidos, novedosos, potencialmente útiles, y comprensibles en los datos (Fayyad, Piatetsky, & Smyth, 1996).

De forma general el proceso involucra los siguientes pasos (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

- **Comprensión del dominio de la aplicación.** En esta fase se identifica el conocimiento previo relevante que debe utilizarse, junto con las metas y los requerimientos del proceso de extracción de conocimiento.
- **Preparación de los datos.** En esta fase se analizan y documentan los datos disponibles y las fuentes de conocimiento donde se involucra los siguientes pasos para completar la fase de preprocesamiento de datos.
 - Limpieza de datos: consta de ciertas operaciones básicas, como eliminar el ruido, tratamiento de valores faltantes o perdidos, redundancia de datos, etc. La información proveniente en un ambiente real suele ser muy incompleta y tener muchos errores, para ello se aplican herramientas que permiten modelar la

información de acuerdo a los requerimientos planteados, para luego esta servir para aplicar los algoritmos de minería de datos.

- Integración de datos: la integración cumple un rol importante en el proceso de minería de datos. Esta herramienta se enfoca en transformar los datos e integrarlos con sus heterogéneos a partir de diferentes fuentes.

2.1.2. Minería de datos

Se define como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas ha transformado el análisis de datos orientados hacia determinadas técnicas especializadas englobadas bajo el nombre de minería de datos o Data Mining (López C. P., 2008).

2.1.3. Minería de datos educacional

Es una disciplina emergente, preocupada por el desarrollo de métodos para explorar los tipos únicos de los datos que provienen de los centros educativos, y el uso de esos métodos para entender mejor a los estudiantes (International Educational Data Mining Society, 2011) .

2.1.4. Soft Computing

El Soft Computing es una rama de la inteligencia artificial centrada en el diseño de sistemas inteligentes capaces de manejar adecuadamente la información incierta, imprecisa y/o incompleta. Esta cualidad permite abordar problemas reales obteniendo soluciones más robustas, manejables y de menor coste que las obtenidas mediante técnicas convencionales (Magdalena, Coso, & Alvarez, 2005).

2.1.5. CRISP-DM

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) estructura el ciclo de vida de un proyecto en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo de la investigación (Chellatamilan, Ravichandran, Suresh, & Kulanthaivel, 2011).

2.1.6. Técnicas descriptivas

Cada una de las tareas presentadas requiere métodos, técnicas o algoritmos para resolverlas. Una tarea puede tener muchos métodos para resolverla y el mismo método (o al menos el mismo tipo de técnica) puede resolver un gran abanico de tareas, dado que la mayoría de las tareas son caras del aprendizaje inductivo (Hernan, Ramirez, & Ferri, 2004).

Algunos de los tipos de técnicas más utilizadas se reseñan en la lista siguiente:

2.1.6.1. Técnicas algebraicas y estadísticas

También denominadas paramétricas, expresan modelos mediante fórmulas, funciones, distribuciones o valores estadísticos como medias, varianzas, etc. Obtienen un patrón a partir de un modelo predeterminado del cual se estiman los coeficientes o parámetros. Son ejemplos de estas técnicas la regresión lineal, regresión logarítmica y logística (Freedman, 2009).

Los modelos en los que el comportamiento de una variable Y se pueden expresar como una función de una variable X se pueden representar mediante $Y = f(X)$, si se considera que la relación f es una función lineal, que las variables explicativas pueden ser N en lugar de una única X y que las relaciones no son exactas, sino más bien aproximaciones, por lo que se debe agregar un término de perturbación aleatoria u que refleje esos factores, la fórmula anterior puede escribirse $Y_i = B_0 + B_1X_{i1} + \dots + B_NX_{iN} + u_i$ y el modelo se denomina regresión lineal.

Cuando la regresión lineal no logra determinar los coeficientes, o el fenómeno en estudio tiene un comportamiento que puede considerarse potencial o logarítmico, se utiliza la regresión logarítmica, que transforma la ecuación anterior aplicando logaritmo a ambos lados de la igualdad.

Un caso de regresión lineal generalizado es la regresión logística que permite evaluar una probabilidad. La variable de respuesta tiene dos o más posibilidades, cada una con su respectiva probabilidad.

2.1.6.2. Técnicas bayesianas

Utilizan el teorema de Bayes para estimar la probabilidad de pertenencia a una clase o grupo. Un ejemplo clásico es el clasificador bayesiano ingenuo o naive Bayes (Winkler, 1972).

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa un atributo y cada arco una dependencia probabilística que expresa la probabilidad condicional de cada atributo dados sus padres. El arco apunta a un atributo dependiente del que está en el origen del arco. La estructura de la red provee información sobre dependencias y también sobre las independencias de un atributo (o conjunto de ellos) de otro u otros. La construcción de una red bayesiana a partir de los datos consta de un proceso de aprendizaje estructural, donde se obtiene la estructura de la red, y un aprendizaje paramétrico en que se obtienen las probabilidades y condicionales de la estructura.

2.1.6.3. Técnicas basadas en conteos de frecuencias y tablas de contingencia

Cuentan la frecuencia en que dos o más sucesos se dan conjuntamente, el algoritmo comienza por pares de sucesos y va incrementando los conjuntos

para los casos en el que las frecuencias conjuntas superen un umbral. El algoritmo “a priori” es un ejemplo de estas técnicas (Agrawal & Srikant, 1994).

2.1.6.4. Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas

Se basan en los algoritmos del tipo “divide y vencerás” como el ID3/C4.5 (Quinlan, 1993) o el CART y los denominados “separa y vencerás” como el CN2 (Clark & Niblett, 1989). Más adelante se profundiza la descripción de estas técnicas.

2.1.6.5. Técnicas relacionales y estructurales

Representan los modelos mediante lenguajes declarativos como los lenguajes lógicos y funcionales. La mayoría de las técnicas de DM trabajan sobre datos en formato atributo-valor (vista minable), para extender el aprendizaje a una representación del conocimiento de forma estructural o relacional, se debe cambiar el lenguaje de representación y usar lógica de primer orden, esta es la idea de las técnicas relacionales (RDM) (Dzeroski & Lavrač, 2001). La programación lógica inductiva (ILP) es una rama del aprendizaje automático en la que la programación lógica se emplea como técnica de representación uniforme de ejemplos, conocimientos de base e hipótesis.

2.1.6.6. Técnicas basadas en redes neuronales artificiales

Son un método de aprendizaje que parte de la presunción de que la capacidad humana de procesar información se debe a la naturaleza biológica del cerebro, y para imitar esta característica se basan en el uso de soportes artificiales similares a los del cerebro. Algunas variantes son las redes multicapas (Laboratory; S. U. S. E.; Widrow, B.; Hoff, E.; of Naval Research, U.S.O.; Corps. U.S.A.S.; Force. U.S.A., ; Navy, U.S., 1960), perceptrón simple (Rosenblatt, 1962), redes de Kohonen (Kohonen, 1982), que permiten realizar clasificación no supervisada, etc.

2.1.6.7. Técnicas basadas en núcleo y máquinas de soporte vectorial

Representan vectorialmente los ejemplos, con un componente real para cada atributo, el vector se suele denominar vector de pesos. El modelo de máquinas de soporte vectorial (SVM) fue presentado en 1992 por Vapnik, Boser y Guyon (Boser, Guyon, & Vapnik, 1992) y descrito en (Cortes & Vapnik, 1995) y (Vapnik, 1998). Intentan maximizar el margen entre los grupos o clases formadas mediante transformaciones llamadas funciones núcleo, que calculan el producto escalar de dos vectores en el espacio de características, es importante la elección de la función núcleo a utilizar, que debe reflejar el conocimiento a priori del problema.

2.1.6.8. Técnicas estocásticas y difusas

Junto con las redes neuronales estas técnicas forman lo que se llama computación flexible. Son ejemplo los métodos evolutivos (Tettamanzi, Tomassini, & Janßen, 2001) y las funciones de lógica difusa, tales como el algoritmo de reglas difusas para generación de reglas de Wang y Mendel (I.X. & J.M., 1992).

Técnicas basadas en casos, en vecindad o distancia: Se basan en las distancias al resto de los elementos, como vecinos más próximos o los algoritmos jerárquicos como Two-step o COBWeb (Fisher, 1987) y los no jerárquicos como k-medias (Moody & Darken, 1989) (MacQueen, 1967).

Dado que son algoritmos que se adaptan a las tareas descriptivas, se analizan en secciones posteriores con el objeto de ser utilizados como primera aproximación al problema planteado.

Todas las tareas (exceptuando quizá las reglas de asociación y correlaciones) y los métodos descritos se centran en la idea del aprendizaje inductivo. El aprendizaje inductivo es un tipo especial de aprendizaje capaz de obtener

reglas o modelos que generalizan o abstraen la evidencia determinada por un conjunto de ejemplos particulares.

El aprendizaje puede ser incremental o no, dependiendo de la forma en que se presentan los datos. Si se considera que los datos pueden ir cambiando por periodos de tiempo (año académico por ejemplo) podría ser interesante utilizar métodos específicos para el aprendizaje incremental, que permite revisar los modelos aprendidos y no tener que realizarlos de nuevo con todos los datos.

2.1.7. Software Keel

Es un programa de código abierto (GPLv3), desarrollado en Java, sirve para evaluar los algoritmos evolutivos para problemas de minería de datos, incluyendo la clasificación de regresiones, clustering, patrones de minería. Éste contiene una gran colección de algoritmos de extracción de conocimientos clásicos, técnicas de Preprocesamiento.

Dispone de una amplia librería de algoritmos de Sistemas Difusos Evolutivos basados en diferentes esquemas: Michigan, Pittsburgh, IRL y GCCL. También dispone de módulos de tratamiento de datos y análisis estadísticos. Tiene una colección de bibliotecas para técnicas de procesamiento previo y posterior para manipulación de datos, métodos de computación suave en conocimiento de extracción y aprendizaje y para proporcionar métodos de investigación científica

Características

KEEL permite al usuario emplear AEs en diferentes tipos de problemas de minería de datos: regresión, clasificación, agrupamiento, asociación, etc.

Los programas implementados se aplican en amplia investigación y metas educativas como: aprendizaje de regla de difusión evolutiva, ajuste de regla de

Mamdani, redes neurales artificiales genéticas, y aprendizaje de sistemas de clasificación (Learning Classifier Systems) (Keel, 2014).

2.1.8. Software Knime

Es un programa informático de código abierto para el desarrollo y ejecución de técnicas de minería de datos, su uso se basa en el diseño de un flujo de ejecución que plasme las distintas etapas de un proyecto de minería de datos (Knime, 2014).

Está desarrollado sobre la plataforma Eclipse y programado en Java, Knime proporciona distintos nodos agrupados en fichas, por ejemplo:

- a) Entrada de datos [IO > Read].
- b) Salida de datos [IO > Write].
- c) Preprocesamiento [Data Manipulation], para filtrar, discretizar, normalizar, filtrar, seleccionar variables.

Minería de datos [Mining], para construir modelos (reglas de asociación, clustering, clasificación, MDS, PCA...).

- d) Salida de resultados [Data Views] para mostrar resultados en pantalla (ya sea de forma textual o gráfica).

2.1.9. Software Rapid Miner

Es un programa informático de código abierto para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores a través de un entorno gráfico. Se usa en investigación educación, capacitación, creación rápida de prototipos y en aplicaciones empresariales (RapidMiner, 2014).

Características

- Desarrollado en Java
- Es multiplataforma

- Representación interna de los procesos de análisis de datos en ficheros XML.
- Permite el desarrollo de programas a través de un lenguaje de script
- Puede usarse de diversas maneras: a través de un GUI, línea de comandos, en batch (lotes), desde otros programas a través de llamadas a sus bibliotecas.
- Incluye gráficos y herramientas de visualización de datos.
- Dispone de un módulo de integración con r y con Weka.

2.2. MARCO REFERENCIAL

La presente investigación busca encontrar un perfil de estudiante graduado mediante técnicas supervisadas o no supervisadas para lo cual se ha revisado las siguientes publicaciones con el fin de dar un mayor aporte al actual trabajo.

En su estudio (KOTSIANTIS, 2007), compara varios métodos de clasificación de minería de datos en un conjunto de datos de 350 estudiantes con información demográfica, resultados de asignaturas y grupos de trabajos obteniendo resultados que determinan la mejor técnica de clasificación para este caso en particular.

La investigación de (Márquez, Romero, & Ventura, 2012) enfoca su estudio en la aplicación de minería de datos utilizando reglas de asociación, para poder determinar que estudiante necesita ayuda extra en el aprendizaje, esta investigación logró descubrir patrones que son útiles para detectar problemas de aprendizaje.

Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación de a la deserción de alumnos universitarios: Esta investigación habla sobre la deserción de estudiantes de una determinada universidad, donde aplica minería de datos para técnicas descriptivas con el fin de explicar el fenómeno de abandono masivo de estudiantes en dicha universidad. Realiza agrupamiento o algoritmos de clusterización mediante el software Rapid miner (Formia, 2012).

Aprendizaje Evolutivo de Reglas Difusas para Descripción de Subgrupos: En esta investigación se utilizan algoritmos combinados, específicamente algoritmos de lógica difusa con algoritmos genéticos para obtener patrones de comportamiento de un determinado grupo de estudiantes. Obteniendo reglas que predicen conductas futuras basadas en lógica difusa y algoritmos genéticos. Determinando así los diferentes subgrupos que comprenden la población a estudiar (González, 2007).

Mejoras en el rendimiento de estudiantes a través de minería de datos educacional: En este trabajo se aplica minería de datos utilizando como métodos *Educational Data Mining (EDM)*, *ID3 Algorithm*, *Knowledge Discovery in Database (KDD)*, el cual con la ayuda de éstos métodos antes descritos se pretende identificar a los estudiantes que carecen de un alto rendimiento y necesitan asesoría estudiantil por parte del docente que imparte educación de alta calidad (Priya & Kumar, 2013).

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

2.3. MATERIALES Y MÉTODOS

2.3.1. Equipos y Materiales.

Los materiales que se utilizarán durante la investigación son los siguientes:

| | Recursos | Cantidad | Valor Unit. | Valor Total |
|-----------------------------|----------------------------------------------------------------|-----------------|--------------------|--------------------|
| Talento humano | ➤ Ing. Jorge Guanín | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ PhD. Amilkar Puris | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ Ing. Raúl Díaz Ocampo | 3 meses. | \$ 0.00 | \$ 0.00 |
| | Investigador y responsable del proyecto ➤ Iván Ponce | 8 meses. | \$ 0.00 | \$ 0.00 |
| Recursos de Software | ➤ SQL Server 2008 | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ Keel | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ Knime | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ RapidMiner | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ LibreOffice | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ Weka | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ SPSS | 8 meses. | \$ 0.00 | \$ 0.00 |
| | ➤ Open proj | 5 hs. | \$ 0.00 | \$ 0.00 |
| Recursos de Hardware | ➤ Computadoras | 1 unid. | \$ 1,150.00 | \$ 1150.00 |
| | ➤ Impresora | 1 unid. | \$ 270.00 | \$ 270.00 |
| | ➤ Pen drive | 2 unid. | \$ 8.00 | \$ 16.00 |
| Recursos otros | ➤ Libros | | \$ 0.00 | \$ 0.00 |
| | ➤ Útiles de oficina | | \$ 20.00 | \$ 20.00 |
| | ➤ Servicio de internet | 8 meses | \$ 30.00 | \$ 240.00 |
| | ➤ Movilización | | \$ 320.00 | \$ 320.00 |
| | ➤ Llamadas | 100 min. | \$ 0.17 | \$ 17.00 |
| | ➤ Anillados | 10 unid. | \$ 1.00 | \$ 10.00 |
| | ➤ Empastados | 6 unid. | \$ 15.00 | \$ 90.00 |
| | ➤ CDs. | 10 unid. | \$ 0.50 | \$ 5.00 |
| TOTAL | | | | \$ 2138.00 |

Tabla 2. Equipos y Materiales

El financiamiento detallado anteriormente corre por cuenta del autor de la investigación.

2.3.2. Métodos y Técnicas utilizados en la Investigación

El método a seguir para el desarrollo de esta investigación son los siguientes:

Método Científico

Se formulará una hipótesis que permitirá encontrar las particularidades comunes de los estudiantes graduados de la UTEQ. La investigación propone la aplicación de un conjunto de reglas que permitirá resolver problemas planteados por la Unidad de Planeamiento Académico en cuanto la predicción de futuros estudiantes graduados; Además de la continua evaluación del modelo obtenido con el fin de mejorar la calidad del mismo, y finalmente se aceptará o rechazará la hipótesis inicial como consecuencia de las pruebas, experiencia acumulada y entrenamiento de los datos.

Método Analítico

Se enfocará como objeto de estudio el conjunto de atributos para poder realizar un análisis minucioso, y poder determinar la relación que existe entre las variables mediante técnicas de selección de atributos, logrando obtener un conjunto idóneo de atributos que servirá para realizar la presente investigación.

Método Inductivo

Este método utiliza el razonamiento para obtener conclusiones que parten de hechos particulares aceptados como válidos, para llegar a conclusiones cuya aplicación sea de carácter general. El método se inicia con un estudio individual de los hechos y se formulan conclusiones universales que se postulan como leyes, principios o fundamentos de una teoría (Bernal, 2010).

A través de este método se concluirá con un conjunto de reglas entregadas por parte de las diferentes técnicas aplicadas, donde permiten identificar el perfil de éxito de estudios de pregrado de un estudiante de la UTEQ.

Método Deductivo

Con la aplicación del método deductivo se podrá identificar los factores que incide en el tiempo que los estudiantes de la UTEQ demoran en culminar sus estudios de Pre Grado.

Método Descriptivo

Al utilizar el método descriptivo se logrará describir de manera detallada los diferentes factores que inciden en el tiempo que los estudiantes culminen sus estudios de Pre Grado de la UTEQ.

2.3.2.1. Tipo de Investigación

Diseño de Investigación

El desarrollo de la presente investigación utiliza la metodología cuasi experimental, debido a que se obtendrán diferentes experimentos para alcanzar un conjunto de reglas de alta precisión y de fácil interpretabilidad que permitan encontrar la relación entre las variables exógenas y las variables endógenas que se obtendrán en la etapa de recopilación y pre procesamiento de datos.

Se realizará un análisis de los diferentes algoritmos propuestos que entreguen resultados con una precisión alta y un número de reglas idóneo, además de que estas sean entendibles y de fácil interpretar. Ayudando a contribuir con el objetivo de la presente investigación.

Ventajas y usos del Método Cuasi Experimental

Esta metodología permite realizar la experimentación de una determinada investigación en un marco de restricciones, tomando en cuenta la falta de

aleatorización. Sin embargo se debe tener en cuenta la limitación en cuanto al valor predictivo en este tipo de estudios., debido a que las relaciones causales son valiosas porque proporcionan el conocimiento de cómo manipular el entorno sistemático (Bernal, 2010).

Pasos del diseño experimental.

Paso 1: Definir las variables en la matriz de conceptualización que se evaluarán en el diseño de la experimentación.

Paso 2: Seleccionar los límites de manipulación de las variables independientes y traducirlos en tratamientos experimentales.

Paso 3: Seleccionar las herramientas o técnicas para medir los resultados de las variables.

Paso 4: Seleccionar un conjunto de datos propicios para realizar los respectivos experimentos.

Paso 5: Obtener un conjunto de vistas minables para aplicar la experimentación.

Paso 6: Aplicar las técnicas investigadas a las diferentes vistas minables y evaluar los diferentes resultados obtenidos.

Paso 7: Seleccionar el conjunto de reglas que mejores resultados arrojaron en base a las variables dependientes.

Paso 8: Evaluar las reglas obtenidas en un ambiente de pruebas.

Paso 9: Analizar las propiedades del grupo.

2.3.2.2. Técnicas de Investigación

Observación Directa

La observación, como técnica de investigación científica, es un proceso riguroso que permite conocer, de forma directa, el objeto de estudio para luego describir y analizar situaciones sobre la realidad estudiada. (Bernal, 2010).

Mediante esta técnica lograremos examinar los resultados que entreguen los diversos algoritmos que se utilizaran como son: la precisión, la desviación estándar, el porcentaje de error, el conjunto de reglas, la base de conocimiento que servirán para el desarrollo de la investigación.

2.3.3. Población y Muestra

La población de la presente investigación, está conformada los estudiantes graduados de la Universidad Técnica Estatal de Quevedo y el conjunto de Docentes que impartieron las diferentes unidades de aprendizajes, en el periodo 2002 - 2008.

El conjunto de Datos que se utilizará para el desarrollo de la presente investigación se la obtuvo a través de Los Sistemas Gestores De Base De Datos de la Institución y de las respectivas secretarías de las diferentes facultades de la UTEQ. Debido a que la población a estudiar no es extensa, se trabajará con el universo de estudio más no con un muestreo.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS

En el presente capítulo se detalla la manera en que se obtuvieron las bases de reglas o conocimiento oculto que permitieron descubrir las tendencias o factores que influyen en la graduación de un estudiante de la Universidad Técnica Estatal de Quevedo. Cabe mencionar que para obtener dichas bases de reglas se aplicaron algoritmos difusos y reglas de clasificación con la aplicación de los softwares de minería de datos KEEL, KNIME y WEKA.

4.1.1. Preparación de la función para extraer los datos.

Dentro del conglomerado de información que posee la base de datos de la universidad, es de interés para la investigación obtener características como carga horaria de docente, información académica, periodos de estudio, tiempo de duración de la carrera, promedios de los respectivos cursos, datos socioeconómicos del estudiante.

De lo anteriormente mencionado, se obtendrá el conjunto de datos que se utilizaría en la presente investigación a través del Sistema Administrador de Base de Datos (MSQL 2008) que cuenta la UTEQ, los cuales fueron extraídos mediante procedimientos almacenados que se crearon. Además, otra parte del conjunto de datos con los que se trabajó fue otorgado por las secretarías de cada Facultad de la Institución Educativa, la cual involucra los años de estudios en que un estudiante tarda en culminar su respectiva carrera.

Se tiene conocimiento que el SIU (sistema informático universitario) es la fuente principal de información en los periodos comprendidos del 2002 al 2008 y que no se utilizaron fuentes de información externas que provean otra información relevante.

La recopilación de datos se reduce a la obtención de los registros de cada tabla del modelo relacional. Para aquello es necesario proteger la integridad del estudiante por lo cual se aisló los datos personales como apellidos y nombres,

cédulas, direcciones, etc. de cada estudiantes, debido a que estas características no son relevantes ni aportan significativamente al modelo.

El anexo # 1 indica el script que se utilizó para obtener el conjunto de datos inicial con el que se logró trabajar durante la investigación. El script fue desarrollado por el administrador del SIU que domina el modelo relacional de la base de datos alojada en el DBMS SQL Server.

4.1.2. Preparación de los datos.

La calidad del conocimiento descubierto no solo depende del algoritmo de minería utilizado, sino también de la calidad de los datos minados. Por ello, después de la recopilación, el siguiente paso en el proceso de KDD es seleccionar y preparar el subconjunto de datos que se van a minar, los cuales constituyen lo que se conoce como vista minable (Hernández, 2004).

Al realizar cualquier tipo de investigación aplicando minería de datos, la primera etapa de desarrollo es la fase de *Pre procesamiento de datos*. Este paso es clave para el desarrollo del proyecto o las posteriores fases y también para obtener buenos resultados, siendo esta fase la más tediosa de las etapas de MD.

Esta fase cubre todas las actividades para construir el conjunto final de los datos que serán utilizados en las herramientas de modelado, incluye la selección de las tablas (o archivos), registros y atributos, así como la transformación y limpieza de los datos. Las operaciones del lenguaje relacional SQL (del inglés Structured Query Language) son un estándar que se adapta perfectamente para esta tarea. En esta etapa se utilizan técnicas de limpieza, transformación y reducción de dimensiones que aseguren la calidad de los datos y su adecuación para ser utilizados por las herramientas de modelado (Formia S. A., 2012).

En el desarrollo de la presente investigación, durante la etapa del pre procesamiento de datos surgieron problemas que se logró contrarrestar con las siguientes técnicas que se menciona en los puntos posteriores.

| Item | Atributo | Descripción | Tipo | Valor |
|------|--------------------------|--------------------------------------------------------------------------------------------------------|------------|-----------------------------|
| 1 | Año de inicio de carrera | Año que ingresó a estudiar la carrera en la universidad | Númerica | 2002 - 2008 |
| 2 | Carrera | Nombre de la carrera de estudio | Categórica | {FI001...FI032} |
| 3 | Periodo | Periodo inicial de la carrera | Númerica | 2002-2008 |
| 4 | Sexo | Género del estudiante | Categórica | {M,F} |
| 5 | edad_ingreso | Edad de ingreso a la Universidad | Númerica | {15..99} |
| 6 | Sostenimiento | Sostenimiento de la institución secundaria (Fiscal, Particular, Fiscomisional, Gubernamental) | Categórica | {F,P,I,G} |
| 7 | Localización | Localización geográfica de la institución secundaria | Categórica | {LOCAL,FUERA_QUEVEDO,OTROS} |
| 8 | promedio_pre | Media de las notas de admisión | Númerica | {1..10} |
| 9 | promedio_1 | Media de las notas del primer curso | Númerica | {1..10} |
| 10 | asis_1 | media de asistencia a clases en el primer año de carrera | Númerica | {1..100} |
| 11 | taprobacion_1 | Tiempo (cursos académicos) que tarda en aprobar el 1er curso | Númerica | {1..3} |
| 12 | IndAprob_1 | Índice de número de materias aprobadas sin suspensión (materias aprobadas/total materias del semestre) | Númerica | {1..5} |
| 13 | PEDAD1 | Rango de edad de docentes menores a 35 años | Númerica | {0..12} |
| 14 | PEDAD2 | Rango de edad de docentes entre 35 y 50 años | Númerica | {0..12} |
| 15 | PEDAD3 | Rango de edad de docentes mayores a 50 años | Númerica | {0..12} |
| 16 | PEXPEDOCEN | | | |
| 17 | PPondEdad | | | |
| 18 | PindAcademico | | | |
| 19 | promedio_2 | Media de las notas del segundo curso | Númerica | {1..10} |
| 20 | asis_2 | Media de asistencia a clases en el segundo año de carrera | Númerica | {1..100} |
| 21 | taprobacion_2 | Tiempo (cursos académicos) que tarda en aprobar el 2do curso | Númerica | {1..3} |
| 22 | IndAprob_2 | Índice de número de materias aprobadas sin suspensión (materias aprobadas/total | Númerica | {1..5} |

| | | | | |
|----|--------------------|--------------------------------------------------------------------------------------------------------|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | materias del semestre) | | |
| 23 | SEDAD1 | Rango de edad de docentes menores a 35 años | Numérica | {1..12} |
| 24 | SEDAD2 | Rango de edad de docentes entre 35 y 50 años | Numérica | {1..12} |
| 25 | SEDAD3 | Rango de edad de docentes mayores a 50 años | Numérica | {1..12} |
| 26 | SEXPEDOCEN | | | |
| 27 | SPondEdad | | | |
| 28 | SindAcademico | | | |
| 29 | promedio_3 | Media de las notas del tercer curso | Numérica | {1..10} |
| 30 | asis_3 | Media de asistencia a clases en el tercer año de carrera | Numérica | {1..100} |
| 31 | taprobacion_3 | Tiempo (cursos académicos) que tarda en aprobar el 3er curso | Numérica | {1..3} |
| 32 | IndAprob_3 | Índice de número de materias aprobadas sin suspensión (materias aprobadas/total materias del semestre) | Numérica | {1..5} |
| 33 | TEDAD1 | Rango de edad de docentes menores a 35 años | Numérica | {1..12} |
| 34 | TEDAD2 | Rango de edad de docentes entre 35 y 50 años | Numérica | {1..12} |
| 35 | TEDAD3 | Rango de edad de docentes mayores a 50 años | Numérica | {1..12} |
| 36 | TEXPEDOCEN | | | |
| 37 | TPondEdad | | | |
| 38 | TIndAcademico | | | |
| 39 | habito_estudios | ¿Posee hábitos de estudios? | Católica | {0,1} |
| 40 | estrategia_estudio | ¿Qué Estrategia utiliza para estudiar? | Católica | {0. No contesta 1. Interpretar consignas de trabajo 2. Concentrarse en la tarea 3. Hacer cuadros comparativos 4. Expresar por escrito u oralmente 5. Resumir 6. Buscar y resumir ideas principales 7. Leer atentamente 8. Memorizar y repetir 9. Establecer relaciones entre conceptos } |
| 41 | num_hijos | ¿Cuántos hijos tiene? | Numérica | {0..6} |
| 42 | tamaño_familia | ¿Número de integrantes de su | Numérica | {0..7} |

| | | | | |
|----|-------------------------|----------------------------------------------------------------------------------------|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | familia? | | |
| 43 | nivel_educacion_familia | ¿De la familia que vive con usted, cuál es el nivel educativo más alto que se alcanzó? | Categórica | {0. Ninguna 1. Primaria Completa 2. Primaria incompleta 3. Secundaria completa 4. Secundaria incompleta 5. Superior completa 6. Superior incompleta 7. Estudios de 4to nivel (Posgrado) } |
| 44 | media_edad_familia | ¿Cuál es la edad promedio de su familia? | Categórica | {1. 30 - 40 2. 41 - 60 3. Más de 60} |
| 45 | estilo_aprendizaje | | Categórica | {1. Visual_activo 2. Aural_reflexivo 3. Read_Lectura 4. Kinésico_pragmático} |
| 46 | ViveCon | ¿Describa con quién vive usted? | Categórica | {1. Soy independiente 2. Solo con Mamá 3. Solo con Papá 4. Ambos Padres 5. Su pareja 6. Otro familiar} |
| 47 | jornada_trabajo | | Categórica | {NO. No trabaja MT. Medio tiempo TC. Tiempo Completo EV . Eventual} |
| 48 | total_matriculas | Total de matrículas en la Universidad | Numérica | {1..15} |
| 49 | Financiamiento | Sostenimiento económico del estudiante. | Categórica | {1. Ayuda familiar para 1 o 2 hijos estudiando 2. Cuenta propia 3. Ayuda familiar para más de 3 hijos estudiando 4. Préstamo o crédito vigente} |
| 50 | Zona | Situación geográfica donde reside el estudiante. | Categórica | {1. Urbana en Quevedo 2. Marginal en Quevedo 3. Fuera de Quevedo con servicios básicos 4. Cualquier zona rural sin servicios básicos} |
| 51 | Ingreso | Ingresos mensuales del estudiante o la familia con la | Categórica | {1. Sobre los \$ USD 400 |

| | | | | |
|----|--------------------|----------------------------------------------------------------------|------------|------------------------------------------------------------------------------------------------------------|
| | | quien vive. | | 2. Desde \$ USD 200 hasta \$ USD 399 3. Desde \$ USD 100 hasta \$ USD 199 4. Menos de \$ USD 99} |
| 52 | Vivienda | Tipo de vivienda en donde reside el estudiante. | Categórica | {1. Propia y de hormigón armado 2. Propia y Madera o mixta 3. Prestada 4. Hipotecada o Arrendada} |
| 53 | estrato_graduacion | Tiempo que demora un estudiante en culminar sus estudios de pregrado | Nominal | {1,2,3} |

Tabla 3. Matriz de Atributos

4.1.2.1. Transformación y selección de atributos

La selección de características es el proceso de identificar y remover la información redundante e irrelevante en la mayor medida posible, permitiendo a los algoritmos trabajar más rápido y con mayor eficacia.

Los atributos seleccionados serán de fundamental importancia para obtener reglas de mejor calidad y agilizar los procesos de minería de datos. Para esta investigación se contó con un número inicial de 53 atributos. Dentro de los atributos que inicialmente se empezó la investigación, se logró reducir esta cantidad de características a un número menor de 15, donde los atributos restantes no brindan mayor beneficio a la investigación.

En la práctica, si bien existe la posibilidad de recurrir al conocimiento del dominio por parte del investigador para realizar el proceso de selección de forma manual, suele tratarse de un problema muy complejo y por lo tanto es necesario utilizar a las técnicas de minería de datos y herramientas que automaticen este proceso.

Un buen conjunto de atributos seleccionados son aquellos que no tienen relación entre sí, pero con respecto a la clase o variable de salida están muy

conexas, por eso el realizar una matriz de correlación fue muy importante para obtener una mejor idea de la relación que tienen los atributos con respecto a los otros.

Por lo general los expertos en el área de minería de datos pueden decidir la eliminación de atributos que considera que no puede aportar en nada al modelo, ya sea porque sus valores son innecesario al objeto de estudio como por ejemplo cédula o nombres de los estudiantes, o también sea porque este atributo tiene demasiado valores perdidos o valores muy inconsistentes que se alejan de la realidad.

Para poder entender qué atributos se descartan para la continuidad de la investigación se ha utilizado una herramienta que permite visualizar de una manera muy rápida la inconsistencia en los datos. Esta herramienta es Weka, pero a gusto del investigador se puede utilizar cualquier herramienta que arroje información estadística como porcentaje de valores nulos, gráficos de barras del comportamiento de los datos, promedios, valores mínimos y máximos, etc.

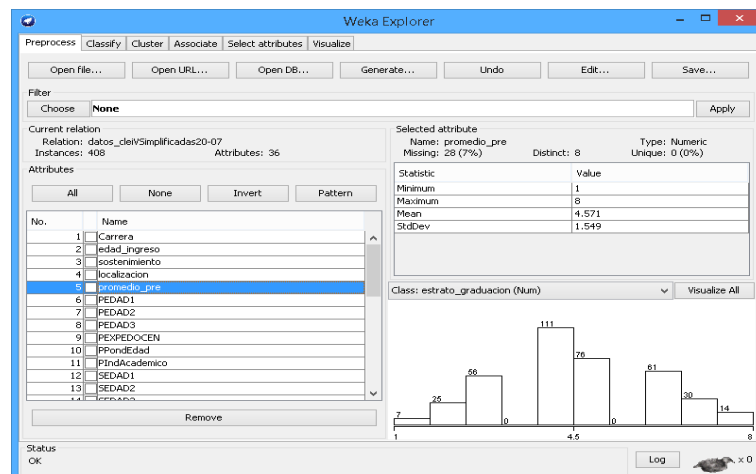


Ilustración 1. Herramienta Weka

Como se puede observar en la imagen anterior, el software Weka permite tener una visión clara de los atributos que pertenecen al conjunto de datos. En este caso, esta herramienta de minería de datos nos brinda información sobre el atributo *promedio_pre*. Este tipo de análisis nos permite identificar la anomalía

e inconsistencia de los datos, presentando información como el promedio, valor mínimo, valor máximo, porcentaje de valores perdidos, desviación estándar.

Además esta herramienta cuenta con un módulo de selección de atributos que permite agilizar y automatizar este proceso.

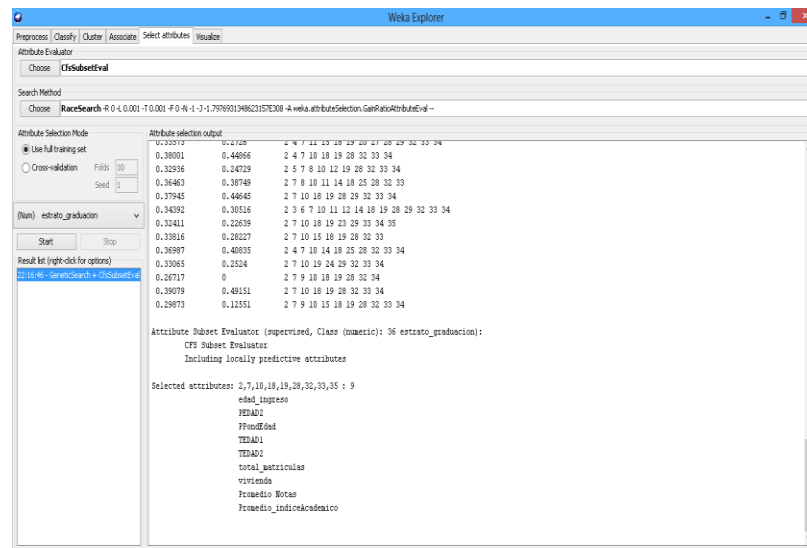


Ilustración 2. Módulo de Selección de Atributos - Weka

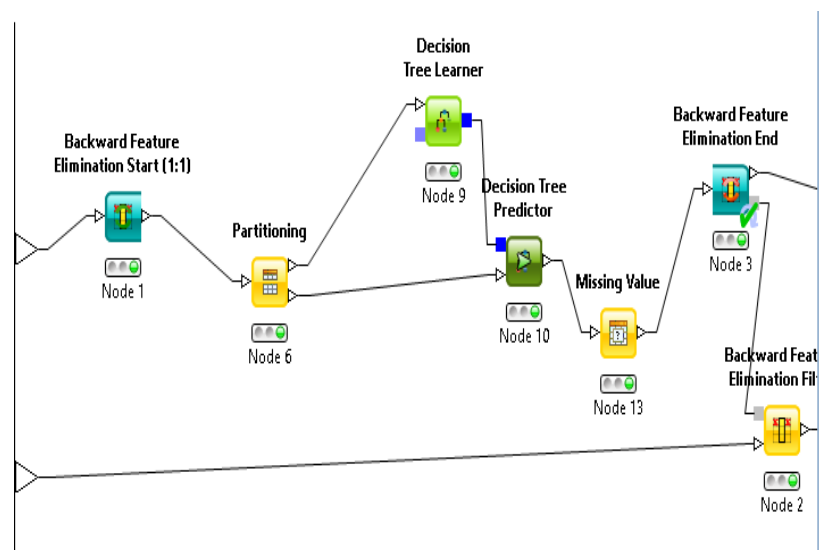
Weka, es una potente herramienta de minería de datos, que entre sus procesos cuenta con selección de atributos. Este módulo incluye la combinación de una búsqueda entre la estimación de la utilidad del atributo y la evaluación respecto a un esquema de aprendizaje específico. Se utilizó los dos algoritmos evaluadores de subconjunto de atributos disponible en el Weka.

Se ejecutó como evaluador al CfsSubsetEval en combinación con el método de búsqueda BestFirst, el cual busca en el espacio de los subconjuntos de atributos utilizando la estrategia greedy hillclimbing con backtracking. La dirección de la búsqueda realizada por el método de búsqueda fue hacia adelante, partiendo del conjunto vacío de atributos.

De igual manera se utilizó el mismo evaluador pero con un método de búsqueda diferente, GeneticSearch el cual realiza una búsqueda genética para la selección de atributos.

El CfsSubsetEval como método evaluador, Calcula un subconjunto de atributos considerando la habilidad individual de cada variables, así como el grado de redundancia entre ellas. Se prefieren los subconjuntos de atributos que estén altamente correlacionados con la clase y tengan baja intercorrelación.

Knime, otra herramienta de minería de datos. Este software contiene nodos que permiten elaborar diagramas de acuerdo a las necesidades del proceso que se esté aplicando. En la siguiente imagen se presenta el diagrama que permite seleccionar atributos a través de una búsqueda hacia atrás. Es decir que del conjunto total de atributos empieza a quitar hasta llegar a obtener los atributos que considera necesario para el modelo.



**Ilustración 3. Diagrama de selección de atributos.
Knime**

La siguiente imagen muestra los resultados obtenidos al haber ejecutado el diagrama antes mencionado. Knime como herramienta de minería de datos presenta varios conjuntos de selección de atributos por cada diagrama realizado. Para seleccionar el conjunto de selección de atributos otorgado por

el software, analizamos los resultados que entrega la aplicación. El porcentaje de error y la cantidad de atributos seleccionados.

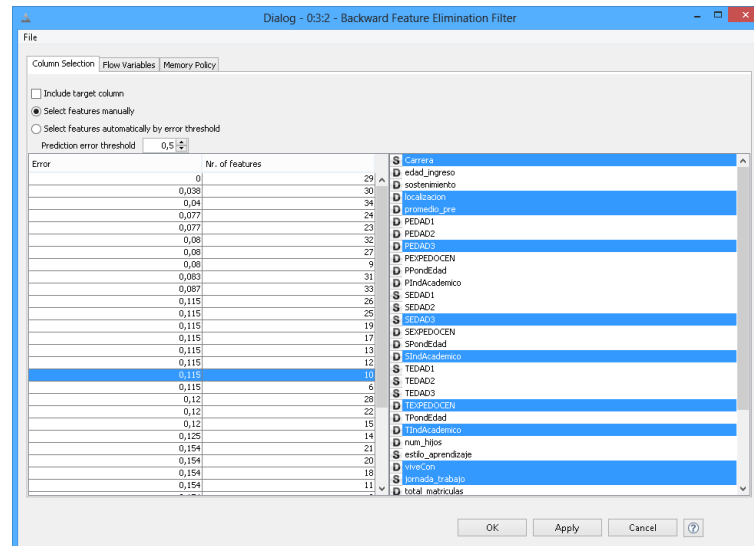


Ilustración 4. Atributos otorgado por el Software

Tipos de algoritmos de selección de atributos.

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

Wrappers. Utilizan el algoritmo de DM que se va a aplicar finalmente a los datos para evaluar la relevancia de los atributos. Parten del razonamiento que el método que se usará finalmente para la predicción debería proveer la mejor estimación de características. Estos algoritmos tienden a ser lentos dado que deben llamar repetidamente al algoritmo elegido. La idea detrás de los modelos wrapper es sencilla: el algoritmo inductivo (aquel que se va a utilizar para la tarea de clasificación) es considerado como una caja negra. Es decir que no es necesario conocer el algoritmo, solamente se necesita conocer su interface (Kohavi & John, 1997). Se ejecuta el algoritmo sobre el conjunto de datos con diferentes conjuntos de características y luego se elige el subconjunto con la mejor evaluación. Luego se evalúa el subconjunto obtenido para un subconjunto de datos no utilizado durante la búsqueda. La figura número 4 ilustra este proceso.

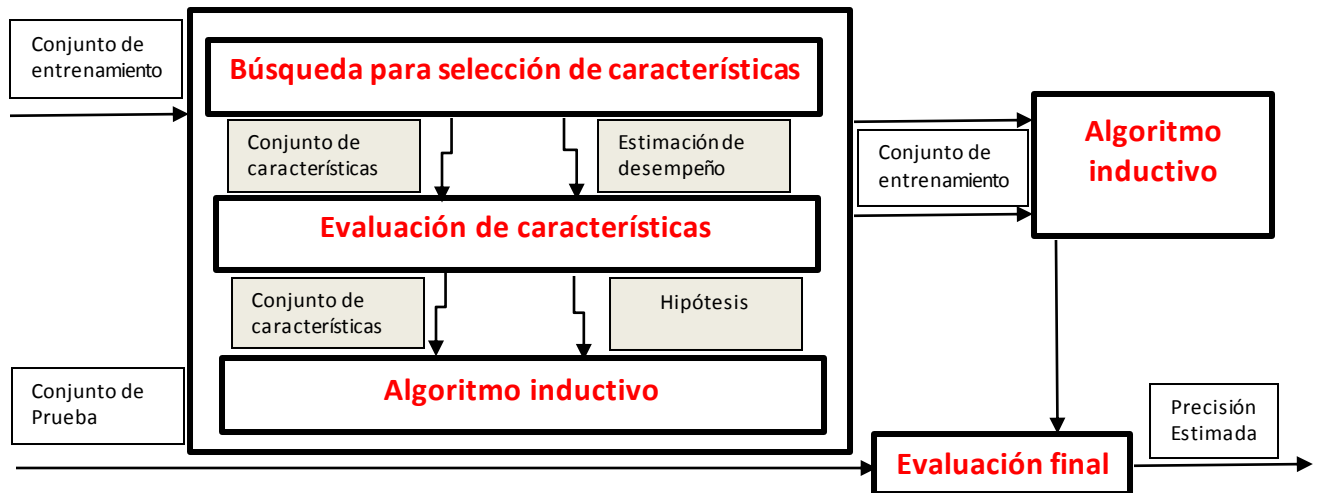


Ilustración 5. Esquema genérico de algoritmo tipo wrapper

Filtros (filters). Seleccionan/evalúan los atributos en forma independiente del algoritmo de aprendizaje (ver figura 5).

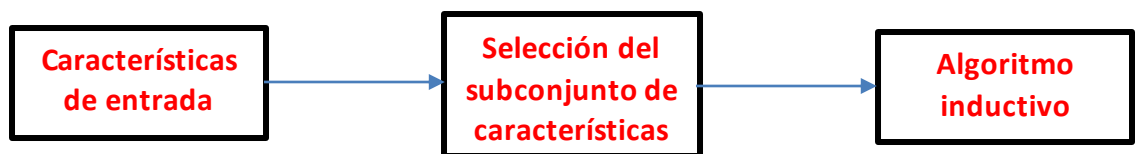


Ilustración 6. Esquema genérico de algoritmo tipo filtro

Híbridos. Usan una combinación de los dos criterios de evaluación en diferentes etapas del proceso de búsqueda. En la siguiente figura se observa el pseudocódigo de un algoritmo tipo híbrido donde implementa los métodos de filtro y wrapper para encontrar un conjunto idóneo de atributos que permitan agilizar el modelo y obtener reglas de mejor calidad.

```

D <- conj. de datos de entrenamiento
Smejor <- 1 Conj. inicial de atributos
Aptitudmejor <- 0

Mientras
  SmejorC = 0
  AptitudmejorC = 0
  Para cada SubConj. Snuevo de atributos de cardinalidad C do
    Aptitudnuevo = eval(Snuevo; D; M1)
    SI Aptitudnuevo > AptitudmejorC Entonces
      AptitudmejorC = Aptitudnuevo
      SmejorC = Smejor
    Fin SI
  Fin Para
  AptitudmejorC = eval(SmejorC; D; M2)
  SI AptitudmejorC > Aptitudmejor Entonces
    Aptitudmejor = AptitudmejorC
    Smejor = SmejorC
  Fin SI
Fin Mientras
Retorna Smejor
  
```

Ilustración 7. Pseudocódigo de un algoritmo híbrido

Los algoritmos que pertenecen al tipo **Filtros** se identifican por evaluar un subconjunto de atributos independientemente del algoritmo de aprendizaje; mientras que los algoritmos de del método **wrappers** evalúan el dicho subconjunto de características utilizando el algoritmo de aprendizaje.

En las posteriores imágenes se explica mediante diagramas el funcionamiento de los modelos de selección de características; el conjunto de atributos seleccionado dependerá de la estrategia de búsqueda y el criterio de parada utilizada. Diferentes criterios de evaluación independientemente del algoritmo de aprendizaje darán lugar a algoritmos filtros. Diferentes algoritmos de aprendizaje generan diferentes algoritmos wrappers.

4.1.2.2. Transformación de atributos

La transformación de atributos depende directamente de la experiencia y conocimiento del investigador. Este decide si un atributo “x” se queda con sus valores originales, o por el contrario en combinación con otro atributo “y” dan lugar a uno nuevo llamado “z”.

Dentro de los atributos tenemos índice Académico de cada año de estudio. Esta característica hace referencia al esfuerzo que hizo un estudiante para aprobar su año de estudio. Se tomó en cuenta las materias aprobadas, las que se quedó supletorio y aquellas que arrastró. Cada categoría de este conjunto se les asignó un valor. Por ejemplo a las materias aprobadas directamente se les dio un valor de (número total de materias vistas en el año dividido para uno), las materias que aprobó en el supletorio se le sumó 0.25 y aquellas que arrastró se le agregó 0.75. Así si un estudiante aprobaba su año de estudio sin supletorio ni arrastre, tendría un índice académico de 1. Pero si no lo hacía directamente, el índice académico se incrementaba dependiendo de cuantas materias tendría en supletorio o arrastre. A este proceso le podemos denominar transformación de atributos.

En esta investigación el investigador considera aplicar dicho concepto a los siguientes atributos, debido a que los nuevos atributos influyen de manera más acertada al comportamiento del modelo, generando reglas con un porcentaje de precisión mayor.

4.1.2.3. Atributos seleccionado de las diferentes vistas minables

| VISTAS MINABLES | ATRIBUTOS SELECCIONADOS |
|-------------------|-------------------------|
| VISTA MINABLE # 1 | Carrera |
| | PEdad2 |
| | TEdad3 |
| | SEdad2 |
| | TotalMatriculas |
| | Vivienda |
| | PromedioNotas |
| VISTA MINABLE # 2 | edad_ingreso |
| | Pedad2 |
| | TEdad1 |
| | TEdad2 |
| | Total_matriculas |
| | Vivienda |
| | PromedioNotas |
| | PromedioIndiceAcademico |
| | PPondEdad |
| VISTA MINABLE # 3 | Carrera |
| | Localizacion |
| | Promedio_pre |
| | SEdad3 |
| | SIndiAcademico |
| | Texpedocen |
| | TIndicAcademico |
| | ViveCon |
| | JornadaTrabajo |
| | PEdad3 |
| VISTA MINABLE # 4 | Promedio_pre |
| | PEdad2 |
| | PPondEdad |
| | SEdad3 |
| | SExpeDocen |
| | SPondEdad |
| | TIndAcademico |
| | ViveCon |

| | |
|--------------------------|-------------------|
| | TotalMatriculas |
| | PromedioNotas |
| VISTA MINABLE # 5 | EdadIngreso |
| | SExpeDocen |
| | SpondEdad |
| | TIndAcademico |
| | NUmHijos |
| | estiloAprendizaje |
| | TotalMatriculas |
| | Zona |
| | Vivienda |
| | Prom_Notas |
| VISTA MINABLE # 6 | CARRERA |
| | FINANCIAMIENTO |
| | PEDAD2 |
| | SEDAD2 |
| | TEDAD3 |
| | TAPROBACION_1 |
| | TAPROBACION_2 |
| | TAPROBACION_3 |
| | ESTEXPDOC |
| | HABITO_ESTUDIOS |
| | ESTRATOAPROB |
| | TAMAÑO_FAMILIA |
| | JORNADA_TRABAJO |
| | TOTAL_MATRICULAS |
| | SOSTENIMIENTO |

Tabla 4. Atributos Seleccionados - Vista Minables

4.1.2.4. Tratamiento de valores faltantes

El inconveniente de valores faltantes es muy común en el rama de análisis inteligente de datos, y por lo tanto es un inconveniente que debe ser resuelto en la etapa del preprocesamiento de datos, es decir antes de aplicar los algoritmos de minería de datos debido a que esto puede perjudicar al funcionamiento del algoritmo, afectando a la calidad de las reglas.

El tipo de algoritmo que el investigador vaya a elegir para el proceso de minería influye directamente en las medidas que se va a tomar para solucionar el problema de valores faltantes.

Como acciones a tomar para contrarrestar este problema citamos las siguientes:

- Ignorar: Si el algoritmo con el que se va a trabajar es de árboles de decisiones, lógica difusa o algoritmos de clusterización y el porcentaje de valores faltantes no es muy alto, no hacer nada es una opción factible. Sin embargo cuando el algoritmo seleccionado está basado en redes neuronales o redes bayesianas ignorar no es opción.
- Reemplazar el valor. Puede ser una solución factible si se aplica la técnica adecuada para el problema planteado. Se puede reemplazar los valores faltantes con un valor fijo, el máximo, el mínimo, el promedio, la frecuencia, o también se puede aplicar algoritmos de imputación para asignar valores.

4.1.2.4.1. Imputación de datos

La imputación de datos es una técnica tradicional y muy conocida para dar solución al problema de tratamiento de datos faltantes. Keel es una herramienta de minería de datos, que entre sus módulos presenta opciones para enmendar este inconveniente.

.En la siguiente figura, Keel nos muestra los algoritmos que tiene implementado para trabajar con valores perdidos.

El objetivo de estos algoritmos es el de devolver un conjunto de datos que tenga el 0% de valores faltantes.

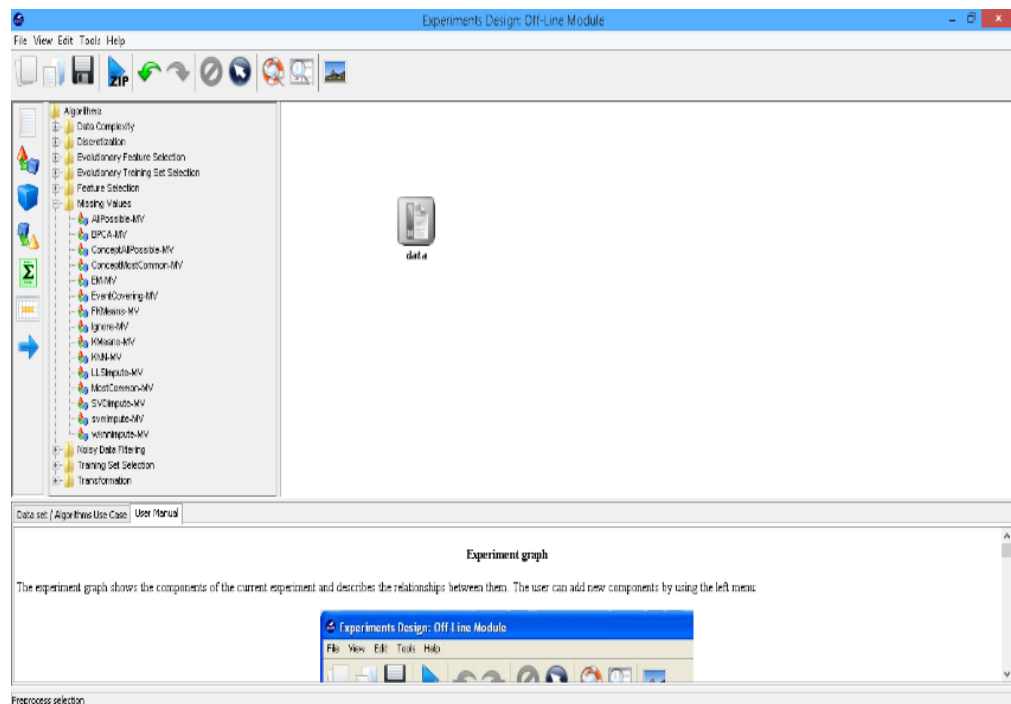


Ilustración 8. Algoritmos Implementados en Keel

4.1.2.5. Técnicas de discretización

En el Pre Procesamiento de la presente investigación se aplicó sub fases como selección de atributos, selección de instancias, métodos de discretización y fusificación, imputación de datos, ausencia de datos.

La aplicación de estas diferentes técnicas y herramientas permitieron obtener varias vistas minables a las cuales se sometieron a diferentes pruebas para obtener la base de reglas.

4.1.3. Extracción de conocimiento

Luego que haya culminado la fase de preprocesamiento de datos como se describe anteriormente, la información resultante es considerada vista minables y está lista para poder ser sometida a las técnicas que el experto en minería de datos considera necesarias para obtener un modelo adecuado que permita finalizar la investigación.

La fase de extracción de conocimiento es la más importante del proceso de MD. Esta etapa presenta una gran variedad de técnicas de las cuales se puede elegir según considere el investigador. La calidad de las reglas va a depender mucho de la técnica elegida.

Aplicar una técnica cualquiera de minería de datos a una de las vistas minables descritas anteriormente, no garantizará el encontrar patrones útiles o novedosos. Es muy común que los algoritmos entreguen resultados con un porcentaje bajo de precisión, esto se debe a que en algunas ocasiones no existe patrones entre los datos, o existe demasiado ruido en la información o inclusive podría ser muy difícil encontrar un patrón, siendo esta la más común. Esto no quiere decir que el algoritmo tiene una baja efectividad, sin embargo no hay que descartar el realizar cambios en las técnicas de minería de datos, que permita mejorar los resultados.

A continuación se expone el proceso de extracción de conocimiento, se detalla las técnicas utilizadas a las diferentes vistas minables obtenidas, y sus resultados, con el fin de demostrar los patrones encontrados que involucren las características comunes de los estudiantes graduados de la UTEQ.

Para el desarrollo de esta fase se utilizó el software keel, e los algoritmos con los que cuenta.

4.1.3.1. Resultados

Con el fin de agotar los caminos que conlleven a buenos resultados se utilizó diez algoritmos a cada una de las vistas minables obtenidas en la fase de preprocesamiento. Los resultados obtenidos al aplicar los respectivos algoritmos de minería de datos no fueron muy alentadores en un principio; se obtuvo reglas con una precisión de 65 a 80 por ciento. Para corregir este inconveniente se aplicó técnicas de sobremuestreo (oversampling), logrando mejorar en una de las vistas minables, obteniendo reglas con una precisión de hasta 90%.

Mientras tanto las vistas minables restantes fueron descartadas de la investigación por su alto porcentaje de error.

La siguiente tabla presenta un resumen de los algoritmos implementados para esta investigación.

| ítem | Algoritmo | Tipo | Enfoque | Referencia |
|------|------------|---------------|----------------------------------|-------------------------------|
| 1 | C45 | Clasificación | Decisions Trees | (Quinlan J. , 1993) |
| 2 | Ripper | Clasificación | Crisp Rule Learning | (Cohen, 1995) |
| 3 | Slipper | Clasificación | Crisp Rule Learning | (Cohen, 1999) |
| 4 | PUBLIC | Clasificación | Decisions Trees | (Rastogi & Shim, 2000) |
| 5 | DT_Oblique | Clasificación | Decisions Trees | (Cantú-Paz & Kamath, 2003) |
| 6 | DT_GA | Clasificación | Decisions Trees | (Carvalho & Freitas, 2004) |
| 7 | PART | Clasificación | Crisp Rule Learning | (Frank & Witten, 1998) |
| 8 | REPSO | Clasificación | Evolutionary Crisp Rule Learning | (Liu, Qin, Shi, & Chen, 2004) |
| 9 | FURIA | Clasificación | Fuzzy Rule Learning | (Hüh & E., 2009) |
| 10 | CHI-RW | Clasificación | Fuzzy Rule Learning | (Chi, Yan, & Pham, 1996) |

Tabla 5. Algoritmos Implementados

En la siguiente tabla se presentan los resultados que se obtuvo aplicando los algoritmos citados.

| Item | Algoritmo | #Reglas | % Precisión pruebas | % Precisión Entrenamiento |
|------|------------|---------|---------------------|---------------------------|
| 1 | C45 | 106 | 86,26 \pm 0.0730 | 94,02 \pm 0.0164 |
| 2 | Ripper | 23 | 91,65 \pm 0.0377 | 97,11 \pm 0.0067 |
| 3 | Slipper | 70 | 92,34 \pm 0.0493 | 98,64 \pm 0.0046 |
| 4 | PUBLIC | 82 | 80,87 \pm 0.0616 | 81,94 \pm 0.0164 |
| 5 | DT_OBLIQUE | 90 | 92,28 \pm 0.0442 | 100,00 \pm 0 |
| 6 | DT_GA | 57 | 78,70 \pm 0.0719 | 79,92 \pm 0.0124 |
| 7 | PART | 63 | 75,82 \pm 0.0545 | 79,20 \pm 0.0758 |
| 8 | REPSO | 6 | 74,22 \pm 0.0871 | 74,55 \pm 0.0277 |
| 9 | FURIA | 46 | 88,89 \pm 0.0511 | 97,31 \pm 0.0092 |
| 10 | CHI-RW | 232 | 73,86 \pm 0.0646 | 97,94 \pm 0.0034 |

Tabla 6. Resultados aplicando los algoritmos implementados.

Se puede observar como los resultados mejoran luego de haber aplicado la técnica de sobremuestreo. El algoritmo que dio una mayor precisión en sus reglas fue el Slipper, sin embargo la cantidad de reglas que arrojó no lo hace el más idóneo para este escenario. El algoritmo Ripper con una precisión de 91.65 % y un número de 23 reglas buenas es el que proporciona la base de reglas resultante como fruto de esta investigación.

4.1.3.2. Base de reglas.

Como resultado de la investigación, se entrega una base de reglas donde involucra las características comunes de los estudiantes graduados.

| ITEM | DESCRIPCION DE LA REGLA |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Si estudia ingeniería en sistemas y se financia por cuenta propia los estudios entonces tarda más de 10 años para graduarse |
| 2 | Si 4 docentes que le dieron clases en primero y 3 docentes en segundo están en una rango de edad entre 35 y 50 años entonces tarda más de 10 años para graduarse |
| 3 | Si tarda 1 año en aprobar el primer año entonces se graduará entre 7 y 10 años |
| 4 | Si tarda 1 año en aprobar el segundo año entonces se graduará entre 7 y 10 años |
| 5 | Si el financiamiento es por Ayuda familiar para 1 o 2 hijos estudiando y 1 docentes de tercero es mayor a 50 años y el estrato de experiencia docente es(EstExpDoc) =2 |
| 6 | Si estudia ing en sistemas y en tercer año no recibió clases de docentes mayores a 50 años entonces se graduará entre 7 y 10 años |
| 7 | Si estudia INGENIERIA EN ADMINISTRACION FINANCIERA, y 2 docentes que le dieron clases en segundo estan en una rango de edad entre 35 y 50 años entonces se graduará entre 7 y 10 años |
| 8 | Si estudia INGENIERIA EN GESTION AMBIENTAL, y su financiamiento es por Ayuda familiar para 1 o 2 hijos estudiando entonces se graduará entre 7 y 10 años |
| 9 | Si en tercer año no recibió clases de docentes mayores a 50 años y el EstratoAprob es 0, entonces se graduará entre 7 y 10 años |
| 10 | Si en tercer año recibió clase de 4 docentes mayores a 50 años y el EstratoAprob es 1, entonces se graduará entre 7 y 10 años |
| 11 | Si el financiamiento es por ayuda familiar para 1 o 2 hijos estudiando y el estratapro=1 y el tamaño de la familia es de 5 entonces se graduará entre 7 y 10 años |
| 12 | Si en segundo año recibió clases con 5 docentes entre 35 y 50 años y el estratoaprob es= 2, entonces se graduará entre 7 y 10 años |
| 13 | Si el financiamiento es por cuenta propia y en primer año recibió clases con 3 docentes entre 35 y 50 años, el tamaño de su familia es de 2, y no trabaja entonces se graduará entre 7 y 10 años |
| 14 | Si en segundo año recibió clases con 4 docentes que tengan edad entre 35 y 50 años, si tiene hábitos de estudio, y su familia está integrada por 4 personas entonces se graduará entre 7 y 10 años |
| 15 | Si estexpdoc= 2 y trabaja de forma eventual, se graduará entre 7 y 10 años |

| | |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 16 | Si en primer año recibió clases de 3 docentes con edad entre 35 y 50 años y en segundo 4 docentes mayores de 50 años, y el EstratoAprob=2 , y su familia está integrada por 4 personas, entonces se graduará entre 7 y 10 años |
| 17 | Si tardó solo 1 año en terminar el tercer año de carrera, entonces se graduará entre 7 y 10 años |
| 18 | Si estudia INGENIERIA EN CONTABILIDAD Y AUDITORIA CPA y el financiamiento de los estudios es por ayuda familiar para 1 o 2 hijos estudiando y la jornada de trabajo es a tiempo completo, entonces se graduará entre 7 y 10 años |
| 19 | si estudia INGENIERIA EN ADMINISTRACION FINANCIERA y se financia por cuenta propia los estudios, y el tamaño de su familia es 6 integrantes entonces se graduará entre 7 y 10 años |
| 20 | Si el total de matrículas es 2 y el sostenimiento es *fiscal entonces se graduará entre 7 y 10 años |
| 21 | Si en segundo año recibió clases de 6 docentes con edades entre 35 y 50 años, el EstratoAprob= 2, el tamaño de la familia es 3, entonces se graduará entre 7 y 10 años |
| 22 | Si estudia INGENIERIA EN GESTION EMPRESARIAL y en primer año recibió clases con 3 docentes y segundo con 2 docentes de edad entre 35 y 50 años, y en tercero con 2 docentes mayores a 50 años y el estexpdoc=1 y estaprob=1, no trabaja y solo se matriculó una vez, entonces se graduará entre 7 y 10 años |
| 23 | Caso contrario se graduará en un tiempo menor o igual a 7 años |

Tabla 7. Bases de reglas

4.1.3.3. Análisis de las Resultados

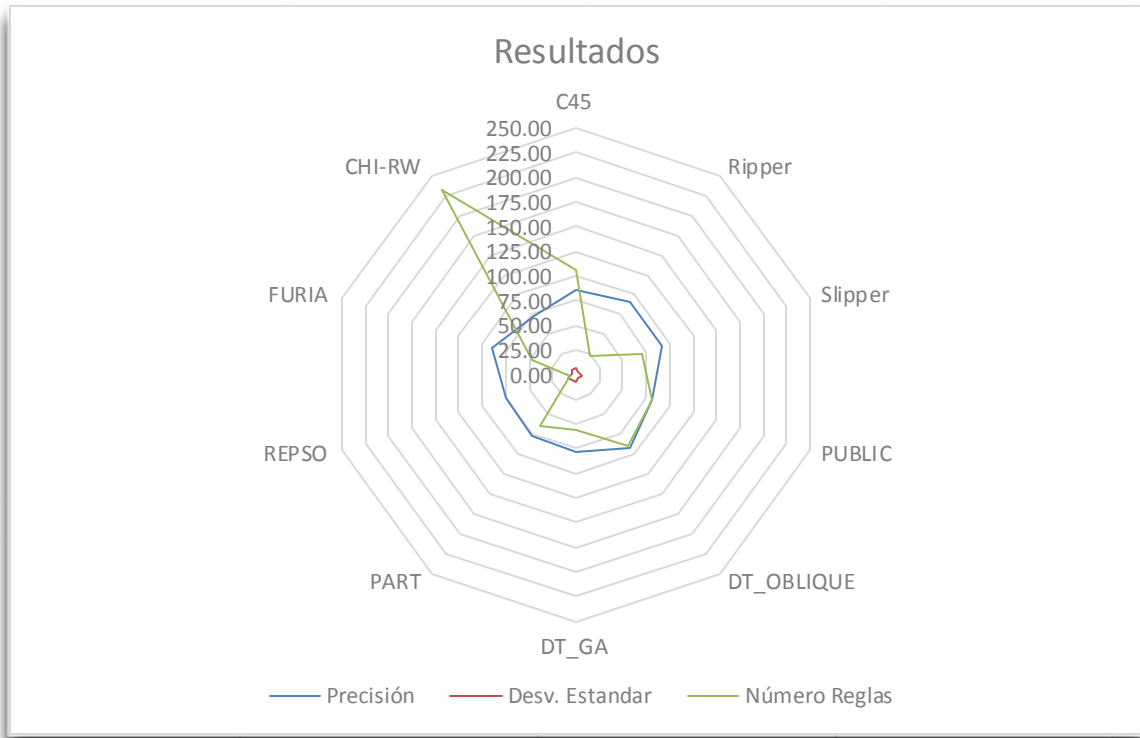


Ilustración 9. Resultados

4.1.3.3.1. Interpretación del gráfico

El gráfico radial nos permite comparar los valores obtenidos de varias series de datos al mismo tiempo. Por su fácil interpretabilidad y además porque permite tener un panorama más claro de los resultados derivados de las respectivas pruebas, se ha elegido como herramienta estadística para describir lo obtenido en la presente investigación.

Se logran ver 3 series de datos. La precisión (línea azul) que indica el acierto que otorgan las reglas de un algoritmo; la Desviación estándar (línea naranja) que hace referencia al valor promedio de la dispersión de la precisión antes mencionada y por último el número de reglas (línea verde) que es el número total de reglas obtenidas a través de aplicar cada uno de los algoritmos antes descritos.

4.1.3.3.1.1. Interpretación del gráfico con respecto a la precisión.

Estas series se van a medir con respecto a los 10 algoritmos aplicados. C45, Ripper, Slipper, Public, Dt_Oblique, Dt_Ga, Part, Repso, Furia, Chi-RW, son algoritmos que se utilizaron durante el desarrollo de la investigación. En una escala de 25 puntos podemos interpretar lo siguiente:

El comportamiento de la precisión bajo cada uno de los algoritmos da a notar de una manera muy visual qué técnica dio mejores resultados. Algoritmos como Chi-RW, Repso, Part y DT_Ga nos devuelve valores con respecto a la precisión entre 70% y 80 %. Estos datos no satisfacen la necesidad de la investigación ni cumple con los objetivos planteado por el autor; quizás debido a la normalización de los datos a analizar, o porque la técnica utilizada por el algoritmo no encontró patrones que modelen el comportamiento de los datos. Esto no quiere decir que se pone en duda la efectividad del algoritmo. Otros Algoritmos como C45, Public presentó valores más elevados en su precisión. 86% y 80% respectivamente. Estos algoritmos basados en árboles de decisiones entregan reglas con un mayor porcentaje de efectividad, acoplándose de mejor manera al comportamiento de la información.

Con estos Resultados ya son candidatos a ser elegidos como los algoritmos que otorgan la base de reglas idónea al interés de la investigación. Aun así se debe tomar en cuenta las otras dos series aun no descritas. Por último tenemos los algoritmos Ripper, Slipper y Dt_Oblique, cuyos resultados de precisión son 91.65%, 92.34%, 92.28%. Estos resultados entregados son realmente muy buenos. Pero en este caso se debe analizar además la desviación estándar y el número de reglas generadas.

4.1.3.3.1.2. Interpretación del gráfico con respecto a la Desviación estándar.

Una desviación estándar de cero indicaría que no existió casualidad en que el resultado de la precisión del algoritmo seleccionado sea muy bueno. Al analizar esta serie de datos encontramos que los algoritmos que despertaron interés al

autor de la presente investigación presentan valores en su desviación estándar muy bajo. El Ripper, Slipper y Dt_Oblique son algoritmos que muestran un valor muy bajo en esta serie de datos; 0.037, 0.049 y 0.044 respectivamente. Afirmando como algoritmos idóneos para solucionar el problema tratado.

Esto no fue fundamental para determinar el algoritmo a seleccionar. Pero dejó muy en claro que los resultados otorgados por el algoritmo no son coincidencias. Dejando como variable fundamental para decidir el número de reglas otorgadas por los algoritmos.

4.1.3.3.1.3. Interpretación del gráfico con respecto al Número de reglas.

Analizar el número de reglas que oferta cada algoritmo fue fundamental para seleccionar el modelo que emule el comportamiento de la información. Los algoritmos que despertaron el interés al autor de la investigación Ripper, Slipper y Dt_Oblique, generaron 23, 70 y 90 reglas respectivamente. Esto nos ayuda a determinar el algoritmo que dio mejores resultados durante el desarrollo de la investigación. Algoritmos como el Slipper y el Dt_Oblique tienen valores de precisión muy altos, y su desviación estándar determina que no son casuales los resultados, sin embargo el alto número de reglas otorgadas hacen que el investigador rechace los algoritmos descritos. Por otro lado el Ripper, algoritmo basado en árboles de decisiones, otorgó un número de reglas conveniente a la línea de la investigación.

El número de reglas otorgadas por los algoritmos fue el factor clave para seleccionar al Ripper como el algoritmo que mejor se adaptó al comportamiento de la base de conocimiento. Obtener un número alto de reglas o caso contrario, un número muy bajo de reglas no es conveniente debido a que las mismas no permiten realizar un análisis correcto, y mucho menos emular el propósito de la investigación.

4.2. DISCUSIÓN

La tabla 7. Base de Reglas. presenta las 23 reglas entregadas por el algoritmo Ripper, siendo el más idóneo para este escenario. Este algoritmo basado en técnicas de clasificación es el que permitió crear mejores reglas para identificar los perfiles de éxito de un estudiante graduado en la UTEQ.

Constratar con otro autor

La base de reglas permite entregar las características comunes que actúan en función de identificar el tiempo de estudio de un alumno de esta institución académica.

Entre los factores endógenos podemos destacar a la carrera como el atributo más dominante en las reglas propuestas. Seguido del tiempo de aprobación de los primeros años y además la experiencia docente.

Como factor exógeno, el que apareció en las reglas otorgadas por el algoritmo fue el financiamiento.

Además es importante analizar las características que no fueron tomadas en cuenta para el diseño de las reglas por parte del algoritmo, como por ejemplo el atributo `jornada_trabajo`, lo que indica que no es excusa un trabajo para lograr graduarse en un plazo menor. Sin embargo no se puede afirmar que así sea, debido a que en los resultados simplemente no presenta reglas con esta característica. Se deja en claro que esto es únicamente una opinión por parte del investigador.

Parece ser que la carrera incide en un gran porcentaje a que un estudiante se logre graduar en un tiempo menor en relación a otro estudiante de otras carreras. El grado de dificultad o la exigencia de cada carrera quizás podría ser la respuesta a esta anomalía. Pero como en la información inicial no se extrajo

información que definan la dificultad de la carrera o la exigencia de la misma no es válida esta versión.

Como resultado de análisis surge que la experiencia docente influye en el tiempo de estudio de un estudiante. A continuación se describirá las reglas más importantes que analizó el autor de la presente investigación. Estas reglas se eligieron para ser descritas porque indica que el tiempo de estudio sería de más de 10 años.

La regla 1 indica que si estudias ingeniería en sistemas el financiamiento es propio tardarás más de 10 años en graduarse. Esta regla realmente está clasificando y analizando que los estudiantes de esta carrera no suelen graduarse inmediatamente.

La regla 2 dice: Si 4 docentes que le dieron clases en primero y 3 docentes en segundo están en un rango de edad entre 35 y 50 años entonces tardará más de 10 años en graduarse. Lo que indica que los docentes más jóvenes no deberían dar clases a los primeros años.

Los años idóneos de estudio deberían de ser menor o igual a siete años, sin embargo estas permiten analizar dos escenarios nada alentadores para el desarrollo académico de la institución que deberían ser examinadas por las autoridades correspondientes.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

Al concluir el presente trabajo de investigación de Estudio de caracterización de alumnos graduados a través de minería de datos educacional y como conocimiento adquirido durante el desarrollo de la misma, se concluye lo siguiente:

- Se identificaron los atributos que aportan información al modelo, que se logró obtener a través de técnicas de minería de datos como selección de atributos, imputación de datos y otras herramientas de DM en la fase de preprocesamiento. La tabla 4. Atributos seleccionados – Vista MInables, indica el conjunto de atributos seleccionados.
- Durante el desarrollo de la presente investigación se estudiaron las técnicas de minería de datos en las diferentes etapas, siendo las técnicas de clasificación las que mejor se adaptaron al modelo para mejorar los resultados obtenidos.
- Mediante el análisis inteligente de datos o minería de datos se logró obtener como resultado de la investigación una base de reglas donde agrupa las características comunes de los estudiantes graduados de la universidad. (ver tabla 7. Base de Reglas).
- Los resultados derivados de esta investigación limitados en cierto punto por restricciones de los datos de nivel complementario hacen prevalecer el objetivo conseguido que es el obtener conocimiento oculto mediante reglas de clasificación.
- El porcentaje de precisión, desviación estándar y el número de reglas, fueron factores claves para determinar el algoritmo que mejor se adaptó al comportamiento de la base de conocimientos. El algoritmo perteneciente a las técnicas de clasificación, Ripper, otorgó resultados

optimistas, demostrando la adaptabilidad de las reglas para identificar el perfil de éxito de un estudiante graduado en a UTEQ.

5.2. RECOMENDACIONES

Luego de la culminación del presente trabajo de investigación y con la experiencia adquirida durante su desarrollo, se recomienda:

- Revisar la base de reglas generadas por la presente investigación por parte de los gestores académicos para que les sirva de apoyo a las decisiones que tengan con respecto a mejorar el índice de graduados de la U.T.E.Q.
- Utilizar la base de reglas obtenidas para comprobar la efectividad de las mismas, y poder evaluarlas en un ambiente real.
- Dar continuidad a la investigación, además de recopilar información de los cursos académicos y extender el abanico de técnicas y herramientas aplicadas para obtener mejores resultados, con una base de reglas de mejor calidad y robustez.

CAPÍTULO VI

BIBLIOGRAFÍA

6.1. LITERATURA CITADA

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association. En *In Proceedings of the 20th International Conference on Very Large Data* (págs. 487 -499). San Francisco.
- Bernal, C. A. (2010). *Metodología de la investigación*. Colombia: PEARSON EDUCACION.
- Biblioteca de Guayaquil. (2011). <http://www.bibliotecadeguayaquil.com/>.
Obtenido de http://www.bibliotecadeguayaquil.com/index.php?option=com_content&view=article&id=221:universidad&catid=43:fechas-historicas&Itemid=137
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. New York.
- Bunge. (2000). *La investigación científica*.
- Canney, E. (2007). Data Mining. 2.
- Chandrani, S., Arpita, G., & Santosh, M. (2011). EXTRACTION AND ANALYSIS OF FACULTY PERFORMANCE OF MANAGEMENT DISCIPLINE FROM STUDENT FEEDBACK USING CLUSTERING AND ASSOCIATION RULE MINING TECHNIQUES. *3rd International Conference on Electronics Computer Technology 4, ICECT 2011 - 2011*, (págs. 94 -96).
- Chellatamilan, Ravichandran, Suresh, & Kulanthaivel. (2011). Effect of Mining educational data to improve Adaptatino of learning in e-learning. 922 - 927.
- 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.
- Clark, P., & Niblett, T. (1989). The cn2 induction algorithm.
- Cortes, C., & Vapnik, V. (1995). *Support-vector networks*.
- Dzeroski, S., & Lavrač, N. (2001). *Relational Data Mining*.
- Fayyad. (1996). *The KDD Process for Extracting Useful Knowledge from Volumes of Data*.
- FAYYAD. (1996). *The KDD Process for Extracting Useful Knowledge from Volumes of Data*.
- Fayyad, Piatetsky, & Smyth. (1996). *From data mining to knowledge discovery*.

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. En *The KDD process for extracting useful knowledge from volumes of data* (págs. 27 - 34).
- Fisher, D. H. (1987). *Knowledge acquisition via incremental conceptual clustering*.
- Formia, S. (2012). *Evaluación de técnicas de Extracción de Conocimiento en Base de Datos y su aplicación a la deserción de alumnos universitarios*. Argentina - La Plata.
- Formia, S. A. (2012). *Evaluación de técnicas de Extracción de conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios*. La Plata.
- Freedman. (2009). *Statistical Models: Theory and Practice*.
- García, P. G. (2007). *Aprendizaje Evolutivo de Reglas Difusas para Descripción de Subgrupos*. Granada.
- González, P. (2007). *Aprendizaje Evolutivo de Reglas Difusas para Descripción de Subgrupos*. España.
- Hernan, O., Ramirez, Q., & Ferri, R. (2004). *Introduccion a la Minería de Datos*. Editorial Pearson.
- Hernández, O. (2004).
- I.X., W., & J.M., M. (1992). Generating fuzzy rules by learning from examples. *IEEE*, 414–1427.
- International Educational Data Mining Society*. (Julio de 2011). Obtenido de <http://www.educationaldatamining.org/>
- Keel*. (01 de Enero de 2014). Obtenido de <http://www.keel.es/>
- 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).
- Knime*. (2014). Obtenido de <https://www.knime.org/>
- Kohavi, R., & John, G. H. (1997). *Wrappers for feature subset selection*.
- Kohonen, T. (1982). *Self-organization and associative memory*.
- KOTSIANTIS. (2007). Supervised Machine Learning: A review of classification Techniques. In *Emerging Artificial Intelligence Applications in computer Engineering*, 249-268.
- Laboratories; S. U. S. E.; Widrow, B.; Hoff, E.; of Naval Research, U.S.O.;Corps. U.S.A.S.; Force. U.S.A., ; Navy, U.S. (1960). *Adaptive switching circuits*.

- Liu. (2012).
- Liu, Y. (2012). Applying data minning in score analysis. *Advances in intelligent and Soft Computing*, 175 - 180.
- López. (s.f.). *Procedimiento y técnicas de selección y orientación*. España: Ediciones TEA.
- López, C. P. (2008). *Minería de datos: técnica y herramientas*. Madrid, España.
- López, P., & Gonzales, S. (2008). *Míneria de Datos Técnicas y Herramientas*. España.
- Machine Learning Group at the University of Waikato. (2013). <http://www.cs.waikato.ac.nz/>. Obtenido de <http://www.cs.waikato.ac.nz/~ml/weka/>
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
- Magdalena, L., Coso, R. d., & Alvarez, A. (Diciembre de 2005). *European Center for Soft Computing*. Obtenido de European Center for Soft Computing: <http://www.softcomputing.es/metaspaces/portal/3/73>
- Márquez, C., Romero, C., & Ventura, S. (2012). Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos. *IEEE*, 109.
- Molina, H. G. (2007). Avance en Informática y Sistemas Computacionales. En H. G. Molina, *CONAIS 2007* (pág. 43). Mexico.
- Moody, J., & Darken, C. J. (1989). Fast learning in networks of locally tuned processing units. *Neural Comput*, 81–294.
- Parack, Zahid, & Merchant. (2012).
- Piatetsky-Shapiro, G. (1991). Knowledge Discovery in Real Databases.
- PRIESCA, P. (s.f.). *Fundación Centro Tecnológico de la Información y la Comunicación (CTIC)* . Obtenido de Fundación Centro Tecnológico de la Información y la Comunicación (CTIC) : <http://www.fundacionctic.org/actualidad-y-divulgacion/noticias/analisis-inteligente-de-datos>
- Priya , K. S., & Kumar, A. (2013). Improving the Student's Performance Using Educational Data Mining. *Int. J. Advanced Networking and Applications* , 1680-1685.
- Quinlan, R. (1993). *Programs for Machine Learning*. San Mateo.
- RapidMiner. (2014). *RapidMiner*. Obtenido de <http://rapidminer.com/>

- 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.
- Romero, & Ventura. (2007).
- Rosenblatt, F. (1962). *Principles of neurodynamics: perceptrons and the theory of brainmechanism*.
- Tettamanzi, A., Tomassini, M., & Janßen, J. (2001). *Soft Computing: Integrating Evolutionary, Neural, and Fuzzy Systems*. Springer.
- Torres, C. A. (2010). Metodología de la investigación. En C. A. Torres, *Metodología de la investigación* (pág. 60).
- VALVERDE. (1992). APLICACIONES DE LA LÓGICA BORROSA. En E. T. RIOS, *APLICACIONES DE LA LÓGICA BORROSA* (pág. 106). MADRID.
- Vapnik, V. (1998). *Adaptive and learning systems for signal processing, communications, and control*.
- Winkler. (1972). *An introduction to Bayesian inference and decision*. New York.

CAPITULO VII

ANEXOS

Anexo # 1. Script del procedimiento almacenado para extraer los datos.

```
use dbsiuPhD
SET NOCOUNT ON
DBCC FREEPROCCACHE
DBCC DROPLEANBUFFERS

declare @semestre as varchar(10)
declare @contCurso as int
declare @ExtraerPromedio as varchar(200)
declare @data_set as varchar(500)
declare @periodo as varchar(10)

--variables para insertar en la tabla temporal
declare @codigo as varchar(15),@sexo as varchar(1),@edad as int,@proPre as
varchar(5),@asispre as varchar(5)
declare @pr1 as varchar(5),@pr2 as varchar(5),@pr3 as varchar(5),@pr4 as
varchar(5),@pr5 as varchar(5),@medpromedios as char(10),@mediaindaproba
char(10)
declare @asi1 as varchar(5),@asi2 as varchar(5),@asi3 as varchar(5),@asi4 as
varchar(5),@asi5 as varchar(5),@sostenimiento as varchar(2)
declare @re1 as varchar(5),@re2 as varchar(5),@re3 as varchar(5),@re4 as
varchar(5),@re5 as varchar(5)
declare @situacion as int,@financiamiento char(2),@zona char(2),@ingreso
char(2),@tenencia char(2),@coefSocioEco char(5)
declare @semesPre as varchar(10),@mens as varchar(60),@semesFacul as varchar(15)
declare @facultad as varchar(5),@auxSemes as varchar(15),@E_situacional as
varchar(15),@Localizacion_Colegio as varchar(50)
declare @PcntAsignaturas1 as varchar(5),@PcntAsignaturas2 as
varchar(5),@PcntAsignaturas3 as varchar(5),@PcntAsignaturas4 as
varchar(5),@PcntAsignaturas5 as varchar(5)
declare @curso_1 as char(3),@visual_1 as char(3),@lectura_1 as
char(3),@auditivo_1 as char(3),@kinesico_1 as char(3)
declare @curso_2 as char(3),@visual_2 as char(3),@lectura_2 as
char(3),@auditivo_2 as char(3),@kinesico_2 as char(3)
declare @curso_3 as char(3),@visual_3 as char(3),@lectura_3 as
char(3),@auditivo_3 as char(3),@kinesico_3 as char(3)
declare @curso_4 as char(3),@visual_4 as char(3),@lectura_4 as
char(3),@auditivo_4 as char(3),@kinesico_4 as char(3)
declare @curso_5 as char(3),@visual_5 as char(3),@lectura_5 as
char(3),@auditivo_5 as char(3),@kinesico_5 as char(3)

declare @habito as char(2),@estrategia as char(2),@hijos as char(2),@tamfamiliar
as char(2),@niveledufamilia as char(2)
declare @promedioedadfamiliar as char(2),@cultura as char(2),@vivecon as
char(2),@estilo_aprendizaje as char(2)
declare @jortrab as char(2),@totalmatr as varchar(10),@estilosdocentes as
int,@horas_labo as varchar(5),@PreguntoEstilo char(5)
declare @eaAuditivo as char(2),@eaLectura as char(2),@eaVisual as
char(2),@eaKinesico as char(2),@eaMultiVA as char(5),@eaMultiVR as
char(5),@eaMultiVK as char(5),@eaMultiAR as char(5),@eaMultiAK as
char(5),@eaMultiRK as char(5),@eaMultiARK as char(5),@eaMultiVAR as
char(5),@eaMultiVRK as char(5),@eaMultiVARK as char(5)
declare @opta as char(3),@basica as char(3),@general as char(3),@inves as
char(3),@profe as char(3)
DECLARE @PC1PR AS CHAR(3), @PC1AG AS CHAR(3), @PC1AU AS CHAR(3),@PC2PR AS
char(3),@PC2AG AS CHAR(3), @PC2AU AS CHAR(3),@PC3PR AS CHAR(3),@PC3AG AS
char(3), @PC3AU AS CHAR(3),@POCAS1 AS CHAR(3),@POCAS2 AS CHAR(3),@POCAS3 AS
char(3),@POTRO1 AS CHAR(3),@POTRO2 AS CHAR(3),@POTRO3 AS CHAR(3),@PEDAD1 AS
char(3),@PEDAD2 AS CHAR(3),@PEDAD3 AS CHAR(3),@PEXPEDOCEN AS CHAR(5),@PPondEdad
as char(5),@PIndAcademico as char(15)
```

```

DECLARE @SC1PR AS CHAR(3), @SC1AG AS CHAR(3), @SC1AU AS char(3), @SC2PR AS
char(3), @SC2AG AS char(3), @SC2AU AS char(3), @SC3PR AS char(3), @SC3AG AS
char(3), @SC3AU AS char(3), @SOCAS1 AS char(3), @SOCAS2 AS char(3), @SOCAS3 AS
char(3), @SOTRO1 AS char(3), @SOTRO2 AS char(3), @SOTRO3 AS char(3), @SEDAD1 AS
char(3), @SEDAD2 AS char(3), @SEDAD3 AS char(3), @SEXPEDOCEN AS char(5), @SPondEdad
as char(5), @SIndAcademico as char(15)
DECLARE @TC1PR AS CHAR(3), @TC1AG AS CHAR(3), @TC1AU AS char(3), @TC2PR AS
char(3), @TC2AG AS char(3), @TC2AU AS char(3), @TC3PR AS char(3), @TC3AG AS
char(3), @TC3AU AS char(3), @TOCAS1 AS char(3), @TOCAS2 AS char(3), @TOCAS3 AS
char(3), @TOTRO1 AS char(3), @TOTRO2 AS char(3), @TOTRO3 AS char(3), @TEDAD1 AS
char(3), @TEDAD2 AS char(3), @TEDAD3 AS char(3), @TEXPEDOCEN AS char(5), @TPondEdad
as char(5), @TIndAcademico as char(15)
DECLARE @CC1PR AS CHAR(3), @CC1AG AS CHAR(3), @CC1AU AS char(3), @CC2PR AS
char(3), @CC2AG AS char(3), @CC2AU AS char(3), @CC3PR AS char(3), @CC3AG AS
char(3), @CC3AU AS char(3), @COCAS1 AS char(3), @COCAS2 AS char(3), @COCAS3 AS
char(3), @COTRO1 AS char(3), @COTRO2 AS char(3), @COTRO3 AS char(3), @CEDAD1 AS
char(3), @CEDAD2 AS char(3), @CEDAD3 AS char(3), @CEXPEDOCEN AS char(5), @CPondEdad
as char(5), @CIndAcademico as char(15)
DECLARE @QC1PR AS CHAR(3), @QC1AG AS CHAR(3), @QC1AU AS char(3), @QC2PR AS
char(3), @QC2AG AS char(3), @QC2AU AS char(3), @QC3PR AS char(3), @QC3AG AS
char(3), @QC3AU AS char(3), @QOCAS1 AS char(3), @QOCAS2 AS char(3), @QOCAS3 AS
char(3), @QOTRO1 AS char(3), @QOTRO2 AS char(3), @QOTRO3 AS char(3), @QEDAD1 AS
char(3), @QEDAD2 AS char(3), @QEDAD3 AS char(3), @QEXPEDOCEN AS char(5), @QPondEdad
as char(5), @QIndAcademico as char(15)
DECLARE @para_1 as char(3), @audi_1 as char(3), @lec_1 as char(3), @visu_1 as
char(3), @kine_1 as char(3)
DECLARE @para_2 as char(3), @audi_2 as char(3), @lec_2 as char(3), @visu_2 as
char(3), @kine_2 as char(3)
DECLARE @para_3 as char(3), @audi_3 as char(3), @lec_3 as char(3), @visu_3 as
char(3), @kine_3 as char(3)
DECLARE @para_4 as char(3), @audi_4 as char(3), @lec_4 as char(3), @visu_4 as
char(3), @kine_4 as char(3)
DECLARE @para_5 as char(3), @audi_5 as char(3), @lec_5 as char(3), @visu_5 as
char(3), @kine_5 as char(3), @cambia_carrera as char(5)
--declare @semes as varchar(15), @curso as int, @paralelo as varchar(2), @semestre
as varchar(15)
--tabla para obtener datos en el dataset

drop table tmpData2003
create table tmpData2003 (codigo varchar(15), carrera varchar(5), sexo
varchar(1), edad_ingreso int, sostenimiento varchar(2), localizacion
varchar(50), promedio_pre varchar(5), asis_pre varchar(5),
promedio_1 varchar(5), asis_1
varchar(5), taprobacion_1 varchar(5), IndAprob_1 varchar(5), PC1PR char(3), PC1AG
char(3), PC1AU char(3), PC2PR char(3), PC2AG char(3), PC2AU char(3), PC3PR
char(3), PC3AG char(3), PC3AU char(3), POCAS1 char(3), POCAS2 char(3), POCAS3
char(3), POTRO1 char(3), POTRO2 char(3), POTRO3 char(3), PEDAD1 char(3), PEDAD2
char(3), PEDAD3 char(3), PEXPEDOCEN char(5), PPondEdad char(5), PIndAcademico
char(15),
paralelo_1 char(3), auditivo_1 as
char(3), lectura_1 as char(3), visual_1 as char(3), kinesico_1 as char(3),
promedio_2 varchar(5), asis_2
varchar(5), taprobacion_2 varchar(5), IndAprob_2 varchar(5), SC1PR char(3), SC1AG
char(3), SC1AU char(3), SC2PR char(3), SC2AG char(3), SC2AU char(3), SC3PR
char(3), SC3AG char(3), SC3AU char(3), SOCAS1 char(3), SOCAS2 char(3), SOCAS3
char(3), SOTRO1 char(3), SOTRO2 char(3), SOTRO3 char(3), SEDAD1 char(3), SEDAD2
char(3), SEDAD3 char(3), SEXPEDOCEN char(5), SPondEdad char(5), SIndAcademico
char(15),
paralelo_2 char(3), auditivo_2 as
char(3), lectura_2 as char(3), visual_2 as char(3), kinesico_2 as char(3),

```

```

                                promedio_3 varchar(5),asis_3
varchar(5),taprobacion_3 varchar(5),IndAprob_3 varchar(5),TC1PR char(3),TC1AG
char(3),TC1AU char(3),TC2PR char(3),TC2AG char(3),TC2AU char(3),TC3PR
char(3),TC3AG char(3),TC3AU char(3),TOCAS1 char(3),TOCAS2 char(3),TOCAS3
char(3),TOTRO1 char(3),TOTRO2 char(3),TOTRO3 char(3),TEDAD1 char(3),TEDAD2
char(3),TEDAD3 char(3),TEXPEDOCEN char(5),TPondEdad char(5),TIndAcademico
char(15),
                                paralelo_3 char(3),auditivo_3 as
char(3),lectura_3 as char(3),visual_3 as char(3),kinesico_3 as char(3),
                                promedio_4 char(5),asis_4
varchar(5),taprobacion_4 varchar(5),IndAprob_4 varchar(5),CC1PR char(3),CC1AG
char(3),CC1AU char(3),CC2PR char(3),CC2AG char(3),CC2AU char(3),CC3PR
char(3),CC3AG char(3),CC3AU char(3),COCAS1 char(3),COCAS2 char(3),COCAS3
char(3),COTRO1 char(3),COTRO2 char(3),COTRO3 char(3),CEDAD1 char(3),CEDAD2
char(3),CEDAD3 char(3),CEXPEDOCEN char(5),CPondEdad char(5),CIndAcademico
char(15),
                                paralelo_4 char(3),auditivo_4 as
char(3),lectura_4 as char(3),visual_4 as char(3),kinesico_4 as char(3),
                                promedio_5 varchar(5),asis_5
varchar(5),taprobacion_5 varchar(5),IndAprob_5 varchar(5),QC1PR char(3),QC1AG
char(3),QC1AU char(3),QC2PR char(3),QC2AG char(3),QC2AU char(3),QC3PR
char(3),QC3AG char(3),QC3AU char(3),QOCAS1 char(3),QOCAS2 char(3),QOCAS3
char(3),QOTRO1 char(3),QOTRO2 char(3),QOTRO3 char(3),QEDAD1 char(3),QEDAD2
char(3),QEDAD3 char(3),QEXPEDOCEN char(5),QPondEdad char(5),QIndAcademico
char(15),
                                paralelo_5 char(3),auditivo_5 as
char(3),lectura_5 as char(3),visual_5 as char(3),kinesico_5 as char(3),
                                habito_estudios
char(2),estrategia_estudio char(2),num_hijos char(2),tamaño_familia
char(2),nivel_educacion_familia char(2),
                                media_edad_familia char(2),tipo_cultura
char(2),estilo_aprendizaje char(2),PreguntoEstilo char(5),viveCon char(2),
                                jornada_trabajo char(2),total_matriculas
varchar(10),financiamiento char(2),zona char(2),ingreso char(2),vivienda char(2),
                                horas_labo varchar(5),
                                auditivo char(3),lectura
char(3),visual char(3),kinesico char(3),eaMultiVA char(5),eaMultiVR
char(5),eaMultiVK char(5),eaMultiAR char(5),eaMultiAK char(5),eaMultiRK
char(5),eaMultiARK char(5),eaMultiVAR char(5),eaMultiVRK char(5),eaMultiVARK
char(5),
                                optativa char(3),basica_especifica
char(3),general char(3),investigacion char(3),profesional char(3),
                                media_promedios
char(10),media_IndAprobacion char(10),cambia_carrera char(5),
                                situacion varchar(15))

set @situacion=0
declare periodosEstudios cursor for (
select * from
--(select distinct lectivo from tbsemestre where codigo_semestre like 'F%' and
substring(codigo_semestre,6,1)='1'and left(lectivo,4)<>2009)as q1)

(select distinct lectivo from tbsemestre where codigo_semestre like 'F%' and
substring(codigo_semestre,6,1)='1' and LEFT(lectivo,4 ) between 2003 and 2003)as
q1)
open periodosEstudios
fetch next from periodosEstudios into @periodo

while @@fetch_status=0
begin
set @contCurso=1

```



```

select
codigo_inst from tbestudio where codigo=@codigo and codigo_inst like 'C%'))
if (@sostenimiento is null) set
@sostenimiento='?'
set @sostenimiento=(case when
@sostenimiento='F' then 1 when @sostenimiento='P' then '2' when
@sostenimiento='?' then '0' else '3' end)
exec pa_DBUAprobacionxResolucionPre_DEA
@codigo,@semesPre,@mens output,@proPre output
exec pa_FALTAS_PRE_DEA @codigo,@asispre
output

-- if isnull(@proPre) set
@proPre='?'
if (@proPre='0' or @proPre is null) set
@proPre='?'---5 toma el valor para indicar que no se sabe el promedio
--if (@asispre=0 or @asispre is null
)set @asispre='?'
--if (@asispre is null) set @asispre='?'

--datos para pimer año
set @auxSemes=@semesFacul
exec pa_Promedios_DEA
@codigo,1,@facultad,0,@pr1 output ,@auxSemes output ,@para_1 output
--print @auxSemes
exec pa_CategoriasDocentes_DEA
@codigo,1,@auxSemes,@PC1PR output, @PC1AG output, @PC1AU output,@PC2PR
output,@PC2AG output
, @PC2AU output,@PC3PR output,@PC3AG
output, @PC3AU output,@POCAS1 output,@POCAS2 output,@POCAS3 output
,@POTRO1 output,@POTRO2 output,@POTRO3
output,@PEDAD1 output,@PEDAD2 output,@PEDAD3 output
,@PEXPEDOCEN output,@PPondEdad
output,@PIndAcademico OUTPUT

if @pr1='?' set @asi1='?'

if @pr1<>'?' exec pa_FALTASFACUL_DEA
@codigo,1,@auxSemes,@asi1 output
--
set @re1=(select count(*) from
tbmatricula where codigo=@codigo and año_semestre=1 and codigo_Semestre like
left(@semesFacul,5)+'%')
if @asi1<>'?' exec
pa_TiempoAprobacion_DEA @codigo,@auxSemes,1,@re1 output else set @re1='?'
if @re1<>'0' exec
pa_Media_Asignaturas_DEA @codigo,@auxSemes,1, @PcntAsignaturas1 output --else set
@PcntAsignaturas1=-1
--print @codigo+ ' - '+@semesFacul + ' -
'+ @pr1 + ' - '+@PcntAsignaturas1
exec dbo.pa_CuentaEstilosDocentes
1,@codigo,@semesFacul,@audi_1 output,@lec_1 output,@visu_1 output,@kine_1
output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK output,@eaMultiAR
output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK output,@eaMultiVAR
output,@eaMultiVRK output,@eaMultiVARK output

--datos de segundo año

exec pa_Promedios_DEA
@codigo,2,@facultad,0,@pr2 output,@auxSemes output ,@para_2 output
exec pa_CategoriasDocentes_DEA
@codigo,2,@auxSemes,@SC1PR output, @SC1AG output, @SC1AU output,@SC2PR

```



```

output,@SC2AG output, @SC2AU output,@SC3PR output,@SC3AG output, @SC3AU
output,@SOCAS1 output,@SOCAS2 output,@SOCAS3 output,@SOTRO1 output,@SOTRO2
output,@SOTRO3 output,@SEDAD1 output,@SEDAD2 output,@SEDAD3 output,@SEXPEDOCEN
output,@SPondEdad output,@SIndAcademico OUTPUT

if @pr2='?' set @asi2='?'
if @pr2<>'?' exec pa_FALTASFACUL_DEA

@codigo,2,@auxSemes,@asi2 output

if @asi2<>'?' exec
pa_TiempoAprobacion_DEA @codigo,@auxSemes,2,@re2 output else set @re2='?'
if @re2<>'0' exec
pa_Media_Asignaturas_DEA @codigo,@auxSemes,2, @PcntAsignaturas2 output --else set
@PcntAsignaturas2=-1
--print @codigo+ ' - '+@semesFacul + ' -
'+ @pr2+ ' - '+@PcntAsignaturas2
exec dbo.pa_CuentaEstilosDocentes
2,@codigo,@semesFacul,@audi_2 output,@lec_2 output,@visu_2 output,@kine_2
output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK output,@eaMultiAR
output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK output,@eaMultiVAR
output,@eaMultiVRK output,@eaMultiVARK output

--datos de tercer año

exec pa_Promedios_DEA
@codigo,3,@facultad,0,@pr3 output,@auxSemes output,@para_3 output
exec pa_CategoriasDocentes_DEA
@codigo,3,@auxSemes,@TC1PR output, @TC1AG output, @TC1AU output,@TC2PR
output,@TC2AG output, @TC2AU output,@TC3PR output,@TC3AG output, @TC3AU
output,@TOCAS1 output,@TOCAS2 output,@TOCAS3 output,@TOTRO1 output,@TOTRO2
output,@TOTRO3 output,@TEDAD1 output,@TEDAD2 output,@TEDAD3 output,@TEXPEDOCEN
output,@TPondEdad output,@TIndAcademico OUTPUT

if @pr3='?' set @asi3='?'
if @pr3<>'?' exec pa_FALTASFACUL_DEA

@codigo,3,@auxSemes,@asi3 output

if @asi3<>'?' exec
pa_TiempoAprobacion_DEA @codigo,@auxSemes,3,@re3 output else set @re3='?'
if @re3<>'0'exec
pa_Media_Asignaturas_DEA @codigo,@auxSemes,3, @PcntAsignaturas3 output --else set
@PcntAsignaturas3=-1
--set @re3=(select count(*) from
tbmatricula where codigo=@codigo and codigo_semestre like left(@semesFacul,5)+'%'
and año_semestre=3)
--print @codigo+ ' - '+@semesFacul + ' -
'+ @pr3+ ' - '+@PcntAsignaturas3
--exec dbo.pa_CuentaEstilosDocentes
1,@codigo,@semesFacul,@auditivo_1 output,@lectura_1 output,@visual_1
output,@kinesico_1 output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK
output,@eaMultiAR output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK
output,@eaMultiVAR output,@eaMultiVRK output,@eaMultiVARK output
exec dbo.pa_CuentaEstilosDocentes
3,@codigo,@semesFacul,@audi_3 output,@lec_3 output,@visu_3 output,@kine_3
output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK output,@eaMultiAR
output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK output,@eaMultiVAR
output,@eaMultiVRK output,@eaMultiVARK output

--datos de cuarto año

```

```

--set
@auxSemes=left(@semesFacul,6)+cast(cast(right(@semesFacul,2)as int)+3 as
varchar(3))

exec pa_Promedios_DEA
@codigo,4,@facultad,0,@pr4 output,@auxSemes output,@para_4 output
exec pa_CategoriasDocentes_DEA
@codigo,4,@auxSemes,@CC1PR output, @CC1AG output, @CC1AU output,@CC2PR
output,@CC2AG output, @CC2AU output,@CC3PR output,@CC3AG output, @CC3AU
output,@COCAS1 output,@COCAS2 output,@COCAS3 output,@COTRO1 output,@COTRO2
output,@COTRO3 output,@CEDAD1 output,@CEDAD2 output,@CEDAD3 output,@CEXPEDOCEN
output,@CPondEdad output ,@CIndAcademico OUTPUT
--set @auxSemes=@semesFacul

if @pr4='?' set @asi4='?'
if @pr4<>'?' exec pa_FALTASFACUL_DEA

@codigo,4,@auxSemes,@asi4 output

if @asi4<>'?' exec
pa_TiempoAprobacion_DEA @codigo,@auxSemes,4,@re4 output else set @re4='?'
if @re4<>'0' exec
pa_Media_Asignaturas_DEA @codigo,@auxSemes,4, @PcntAsignaturas4 output --else set
@PcntAsignaturas4=-1

--set @re4=(select count(*) from
tbmatricula where codigo=@codigo and codigo_semestre like left(@semesFacul,5)+'%'
and año_semestre=4)

--print @codigo+ ' - '+@semesFacul + ' -
'+ @pr4+ ' - '+@PcntAsignaturas4

exec dbo.pa_CuentaEstilosDocentes
4,@codigo,@semesFacul,@audi_4 output,@lec_4 output,@visu_4 output,@kine_4
output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK output,@eaMultiAR
output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK output,@eaMultiVAR
output,@eaMultiVRK output,@eaMultiVARK output

--datos de quinto año
--set
@auxSemes=left(@semesFacul,6)+cast(cast(right(@semesFacul,2)as int)+4 as
varchar(3))

exec pa_Promedios_DEA
@codigo,5,@facultad,0,@pr5 output,@auxSemes output,@para_5 output
exec pa_CategoriasDocentes_DEA
@codigo,5,@auxSemes,@QC1PR output, @QC1AG output, @QC1AU output,@QC2PR
output,@QC2AG output, @QC2AU output,@QC3PR output,@QC3AG output, @QC3AU
output,@QOCAS1 output,@QOCAS2 output,@QOCAS3 output,@QOTRO1 output,@QOTRO2
output,@QOTRO3 output,@QEDAD1 output,@QEDAD2 output,@QEDAD3 output,@QEXPEDOCEN
output,@QPondEdad output,@QIndAcademico OUTPUT
--set @auxSemes=@semesFacul

if @pr5='?' set @asi5='?'
if @pr5<>'?' exec pa_FALTASFACUL_DEA

@codigo,5,@auxSemes,@asi5 output --else
if @asi5<>'?' exec
pa_TiempoAprobacion_DEA @codigo,@auxSemes,5,@re5 output else set @re5='?'
if @re5<>'0' exec
pa_Media_Asignaturas_DEA @codigo,@auxSemes,5, @PcntAsignaturas5 output --else set
@PcntAsignaturas5=-1

exec dbo.pa_CuentaEstilosDocentes
5,@codigo,@semesFacul,@audi_5 output,@lec_5 output,@visu_5 output,@kine_5
output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK output,@eaMultiAR
output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK output,@eaMultiVAR
output,@eaMultiVRK output,@eaMultiVARK output

```

```

--print @codigo+ ' - '+@semesFacul + ' -
'+ @pr5+ ' - '+@PcntAsignaturas5
--set @re5=(select count(*) from
tbmatricula where codigo=@codigo and codigo_semestre like left(@semesFacul,5)+'%'
and año_semestre=5)
set @eaAuditivo=0
set @eaLectura=0
set @eaVisual=0
set @eaKinesico=0
--print @semesFacul

--exec dbo.pa_CuentaEstilosDocentes
0,@codigo,@semesFacul,@eaAuditivo output,@eaLectura output,@eaVisual
output,@eaKinesico output,@eaMultiVA output,@eaMultiVR output,@eaMultiVK
output,@eaMultiAR output,@eaMultiAK output,@eaMultiRK output,@eaMultiARK
output,@eaMultiVAR output,@eaMultiVRK output,@eaMultiVARK output

exec pa_AreaFormacionCarreras_DEA
@semesFacul, @opta output,@basica output,@general output, @inves output,@profe
output

--if @eaAuditivo ='-1' set @eaAuditivo
='?'
--if @eaVisual ='-1' set @eaVisual = '?'
--if @eaLectura ='-1' set @eaLectura
='?'
--if @eaKinesico ='-1' set @eaKinesico
='?'

exec dbo.pa_EstiloAprendizaje_DEA
@codigo ,@estilo_aprendizaje output,@PreguntoEstilo output

exec pa_complementos_DEA @codigo,@habito
output,@estrategia output ,@hijos output,@tamfamiliar output,
@niveledufamilia
output,@promedioedadfamiliar output,@cultura output,@vivecon output

exec pa_JornadaTrabajo_DEA
@codigo,@jortrab output
--exec dbo.pa_numeromatriculas_DEA
@codigo,@semesFacul,@totalmatr output

declare @contmtr as integer,@acumtr as
float
set @contmtr=0
set @acumtr=0
if @re1<>'?' set @acumtr=(CAST(@re1 AS
float)) else set @contmtr=@contmtr+1
if @re2<>'?' set
@acumtr=@acumtr+(CAST(@re2 AS float)) else set @contmtr=@contmtr+1
if @re3<>'?' set
@acumtr=@acumtr+(CAST(@re3 AS float)) else set @contmtr=@contmtr+1
if @re4<>'?' set
@acumtr=@acumtr+(CAST(@re4 AS float)) else set @contmtr=@contmtr+1
if @re5<>'?' set
@acumtr=@acumtr+(CAST(@re5 AS float)) else set @contmtr=@contmtr+1
set @contmtr=5-@contmtr

```

```

@acumtr=@acumtr/@contmtr
varchar(10))

@codigo,@horas_labo output

--obtiene informacion del financiamiento de los estudios
if exists(select * from tbfichasocioeconomica where codigo=@codigo)
    select @financiamiento=(case when financiamiento=5 then 4 else
financiamiento end) from tbfichasocioeconomica where codigo=@codigo
else set @financiamiento=3
--obtiene informacion de la zona donde vive
if exists(select * from tbfichasocioeconomica where codigo=@codigo)
    select @zona=(case when zona_residencial=5 then 4 else zona_residencial
end) from tbfichasocioeconomica where codigo=@codigo
else set @zona=2
--obtiene informacion del ingreso de la familia con quien vive el alumno
if exists(select * from tbfichasocioeconomica where codigo=@codigo)
    select @ingreso=(case when ingreso_familiar =5 then 4 else
ingreso_familiar end) from tbfichasocioeconomica where codigo=@codigo
else set @ingreso=2
--obtiene informacion del tipo de vivienda donde vive el alumno
if exists(select * from tbfichasocioeconomica where codigo=@codigo)
    select @tenencia=(case when tenencia_vivienda=5 then 4 else
tenencia_vivienda end)from tbfichasocioeconomica where codigo=@codigo
    else set @tenencia=4
    --set @coefScocioEco=cast(cast((cast(@financiamiento as
float)+cast(@zona as float)+cast(@ingreso as float)+cast(@tenencia as float))/4
as float) as char(5))

if @pr1<>'?' set @situacion=@situacion+1
if @pr2<>'?' set @situacion=@situacion+1
if @pr3<>'?' set @situacion=@situacion+1
if @pr4<>'?' set @situacion=@situacion+1
if @pr5<>'?' set @situacion=@situacion+1

set @E_situacional=(case when @situacion
=5 then '1'--'TERMINO'
'2'--'PENDIENTE'

when @situacion <5 and @situacion>0 then

when @situacion =0 then '3'--'ABANDONO'
else '0'
end)

if @pr1='?' set @pr1='0'
if @pr2='?' set @pr2='0'
if @pr3='?' set @pr3='0'
if @pr4='?' set @pr4='0'
if @pr5='?' set @pr5='0'

if @PcntAsignaturas1='?' set
if @PcntAsignaturas2='?' set

```

```

@PcntAsignaturas3='0'

@PcntAsignaturas4='0'

@PcntAsignaturas5='0'

CHAR(10))

if @PcntAsignaturas3='?' set

if @PcntAsignaturas4='?' set

if @PcntAsignaturas5='?' set

/*declare @auxpro as float
set @auxpro=(@pr1+@pr2+@pr3+@pr4+@pr5)/5
set @medpromedios=CAST(@auxpro as

print @pr1
print @pr2
print @pr3
print @pr4
print @pr5*/
set @medpromedios=cast(
((cast(isnull(@pr1,0) as float)+cast(isnull(@pr2,0) as float)+cast(isnull(@pr3,0)
as float)+cast(isnull(@pr4,0) as float)+cast(isnull(@pr5,0) as float))/5 ) as
char(10))

set @mediaindaproba= CAST( ((ISNULL(
CAST(@PcntAsignaturas1 as float),0)+ISNULL( CAST(@PcntAsignaturas2 as
float),0)+ISNULL( CAST(@PcntAsignaturas3 as float),0)+ISNULL(
CAST(@PcntAsignaturas4 as float),0)+ISNULL( CAST(@PcntAsignaturas5 as
float),0))/5) AS char (10))

exec pa_CambiaCarrea @codigo,

@cambia_carrera output

-- *****
INSERCIÓN DE DATOS EN LA TABLA *****

INSERT INTO tmpData2003 values(
@codigo,@semesFacul,@sexo,@edad,@sostenimiento,@Localizacion_Colegio,@proPre
,@asispre,

@pr1,@asi1,@re1,@PcntAsignaturas1,@PC1PR , @PC1AG , @PC1AU ,@PC2PR
,@PC2AG , @PC2AU ,@PC3PR ,@PC3AG , @PC3AU ,@POCAS1 ,@POCAS2 ,@POCAS3 ,@POTRO1
,@POTRO2 ,@POTRO3 ,@PEDAD1 ,@PEDAD2 ,@PEDAD3 ,@PEXPEDOCEN
,@PPondEdad,@PIndAcademico,

@para_1,@audi_1,@lec_1,@visu_1,@kine_1,

@pr2,@asi2,@re2,@PcntAsignaturas2,@SC1PR , @SC1AG , @SC1AU ,@SC2PR
,@SC2AG , @SC2AU ,@SC3PR ,@SC3AG , @SC3AU ,@SOCAS1 ,@SOCAS2 ,@SOCAS3 ,@SOTRO1
,@SOTRO2 ,@SOTRO3 ,@SEDAD1 ,@SEDAD2 ,@SEDAD3 ,@SEXPEDOCEN
,@SPondEdad,@SIndAcademico,

@para_2,@audi_2,@lec_2,@visu_2,@kine_2,

@pr3,@asi3,@re3,@PcntAsignaturas3,@TC1PR , @TC1AG , @TC1AU ,@TC2PR
,@TC2AG , @TC2AU ,@TC3PR ,@TC3AG , @TC3AU ,@TOCAS1 ,@TOCAS2 ,@TOCAS3 ,@TOTRO1
,@TOTRO2 ,@TOTRO3 ,@TEDAD1 ,@TEDAD2 ,@TEDAD3 ,@TEXPEDOCEN
,@TPondEdad,@TIndAcademico,

@para_3,@audi_3,@lec_3,@visu_3,@kine_3,

@pr4,@asi4,@re4,@PcntAsignaturas4,@CC1PR , @CC1AG , @CC1AU ,@CC2PR
,@CC2AG , @CC2AU ,@CC3PR ,@CC3AG , @CC3AU ,@COCAS1 ,@COCAS2 ,@COCAS3 ,@COTRO1
,@COTRO2 ,@COTRO3 ,@CEDAD1 ,@CEDAD2 ,@CEDAD3 ,@CEXPEDOCEN
,@CPondEdad,@CIndAcademico,

@para_4,@audi_4,@lec_4,@visu_4,@kine_4,

```

```

        @pr5,@asi5,@re5,@PcntAsignaturas5,@QC1PR , @QC1AG , @QC1AU ,@QC2PR
,@QC2AG , @QC2AU ,@QC3PR ,@QC3AG , @QC3AU ,@QOCAS1 ,@QOCAS2 ,@QOCAS3 ,@QOTRO1
,@QOTRO2 ,@QOTRO3 ,@QEDAD1 ,@QEDAD2 ,@QEDAD3 ,@QEXPEDOCEN
,@QPondEdad,@QIndAcademico,

        @para_5,@audi_5,@lec_5,@visu_5,@kine_5,

        @habito,@estrategia,@hijos,@tamfamiliar,@niveledufamilia,

        @promedioedadfamiliar,@cultura,@estilo_aprendizaje,@PreguntoEstilo,@viveco
n,

        @jortrab,@totalmatr, @financiamiento,@zona,@ingreso,@tenencia

        ,@horas_labo,

        cast(@eaAuditivo as CHAR(3)),cast(@eaLectura as
CHAR(3)),cast(@eaVisual as CHAR(3)),cast(@eaKinesico as CHAR(3)),@eaMultiVA
,@eaMultiVR ,@eaMultiVK ,@eaMultiAR ,@eaMultiAK,@eaMultiRK
,@eaMultiARK,@eaMultiVAR,@eaMultiVRK ,@eaMultiVARK,

        @opta,@basica,@general,@inves,@profe,

        @medpromedios,

        @mediaindaproba,@cambia_carrera,

        @E_situacional)

        set @pr1=0 set @pr2=0 set @pr3=0 set
@pr4=0 set @pr5=0 set @proPre=0
        set @asi1=0 set @asi2=0 set @asi3=0 set
@asi4=0 set @asi5=0 set @asispre=0
        set @re1='' set @re2='' set @re3='' set
@re4='' set @re5='' set @situacion=0 set @sostenimiento=0
        set @habito='?' set @estrategia='?' set
@hijos='?' set @tamfamiliar='?' set @niveledufamilia='?'
        set @promedioedadfamiliar='?' set
@cultura='?' set @estilo_aprendizaje='?' set @vivecon='?'

        fetch next from alumnos into
@codigo,@sexo,@edad,@semesFacul

        end
        close alumnos
        deallocate alumnos
        --print cast(@contCurso as varchar(2))+ ' --- '+@periodo
        set @contCurso=@contCurso+1
        fetch next from periodosEstudios into @periodo
    end

    --set @contCurso=1--inicia en 1 para volver a evaluar con el
siguiente periodo y desde 1 curso

    end
close periodosEstudios
deallocate periodosEstudios

```

```
select * from tmpData2003 INNER JOIN TbAreaFormCarrera
on tmpData2003.carrera =TbAreaFormCarrera.idcarrera inner join dsgeneraluteq2015
on tmpData2003.codigo =dsgeneraluteq2015.codigo
--where situacion='termino'
```