



UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA
FACULTAD DE INGENIERÍA
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO
MAESTRÍA EN INGENIERÍA EN COMPUTACIÓN

DESARROLLO DE SISTEMAS DE PREDICCIÓN POR CLASIFICACIÓN PARA
ESTIMAR EL DESEMPEÑO ACADÉMICO DE
ALUMNOS DE NUEVO INGRESO

PRESENTA

LUIS ALBERTO LOZOYA TREVIZO

CHIHUAHUA, CHIH.

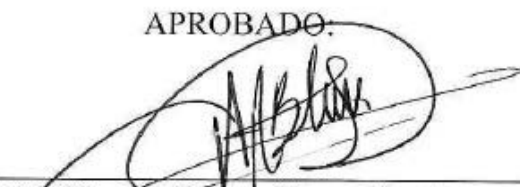
DICIEMBRE /2016

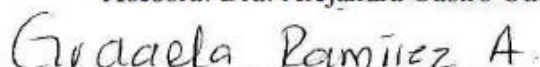


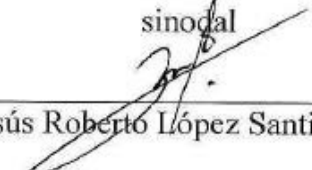
UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA
FACULTAD DE INGENIERÍA
MAESTRÍA EN INGENIERÍA EN COMPUTACIÓN

DESARROLLO DE SISTEMAS DE PREDICCIÓN POR CLASIFICACIÓN
PARA ESTIMAR EL DESEMPEÑO ACADÉMICO DE
ALUMNOS DE NUEVO INGRESO

APROBADO:


Dr. Ricardo Blanco Vega, director
Asesora: Dra. Alejandra Castro Cuesta


Dra. Graciela María de Jesús Ramírez Alonso,
sinodal


M.I. Jesús Roberto López Santillán, sinodal

CHIHUAHUA, CHIH.

15/12/2016

Derechos reservados
© Luis Alberto Lozoya Trevizo
Calle Lucio Cabañas No. 1002, Col. Div. del Norte,
Chihuahua, Chihuahua
2016

Copyright ©

por

Luis Alberto Lozoya Trevizo

2016

6 de diciembre de 2016



ING. LUIS ALBERTO LOZOYA TREVIZO

Presente

En atención a su solicitud relativa al trabajo de tesis para obtener el grado de Maestro en Ingeniería en Computación, nos es grato transcribirle el tema aprobado por esta Dirección, propuesto y dirigido por el director **Dr. Ricardo Blanco Vega** para que lo desarrolle como tesis, con el título: **“Desarrollo de sistemas de predicción por clasificación para estimar el desempeño académico de alumnos de nuevo ingreso”**.

ÍNDICE

RESUMEN

ÍNDICE DE CONTENIDO

CAPÍTULO I. INTRODUCCIÓN

- 1.2 Antecedentes
- 1.3 Problema de investigación
- 1.4 Pregunta de investigación
- 1.5 Objetivos
- 1.6 Aportación práctica
- 1.7 Objeto de estudio
- 1.8 Ideas a defender
- 1.9 Hipótesis

CAPÍTULO II. JUSTIFICACIÓN

CAPÍTULO III. FUNDAMENTACIÓN TEÓRICA

- 3.1 Desempeño académico
- 3.2 Bases de datos
- 3.3 Minería de datos
- 3.4 Netbeans
- 3.5 Autolt

Facultad de Ingeniería

Circuito No.1, Campus Universitario 2

Chihuahua, Chih. C.P. 31125

Tel. (614) 442-95-00 www.fing.uach.mx



CAPÍTULO IV. VALIDACIÓN EXPERIMENTAL

- 4.1 Comprensión del negocio
- 4.2 Comprensión de los datos
- 4.3 Preparación de los datos
- 4.4 Modelado
- 4.5 Evaluación
- 4.6 Implementación

CAPÍTULO V. CONCLUSIÓN

REFERENCIAS

Solicitamos a Usted tomar nota de que el título del trabajo se imprima en lugar visible de los ejemplares de las tesis.

ATENTAMENTE
"naturam subiecit aliis"

EL DIRECTOR

M.I. JAVIER GONZÁLEZ CANTÚ

FACULTAD DE
INGENIERÍA
UACH.



DIRECCIÓN

EL SECRETARIO DE INVESTIGACIÓN
Y POSGRADO

DR. FERNANDO RAFAEL ASTORGA
BUSTILLOS

RESUMEN

El objetivo principal de los estudiantes al entrar en instituciones de nivel superior es el de convertirse en profesionistas, sin embargo no todos logran terminar sus estudios. Esto es un hecho preocupante para las instituciones de nivel superior, ya que buscan mejorar su formación educativa para que los alumnos terminen con sus estudios. Para la institución representa un problema el identificar a los alumnos de nuevo ingreso que puedan tener un desempeño académico bajo y por lo tanto no pueden tomar medidas preventivas que ayuden a los alumnos con sus estudios, es por eso que mediante el uso de la minería de datos se desarrollará un sistema de predicción con algoritmos de clasificación que identifique a los alumnos con probabilidades de desempeño académico bajo en su primer cuatrimestre y así los directivos puedan tomar las medidas necesarias para que los alumnos continúen con sus estudios. Este sistema de predicción se llevará a cabo en la Universidad Tecnológica de Chihuahua UTCH donde se analizarán los datos obtenidos de la institución.



ÍNDICE DE CONTENIDO

RESUMEN	iv
ÍNDICE DE CONTENIDO	v
CAPÍTULO I. INTRODUCCIÓN.....	8
1.2 Antecedentes	8
1.3 Problema de Investigación.....	12
1.4 Pregunta de Investigación.....	12
1.5 Objetivos.....	12
1.6 Aportación Práctica.....	13
1.7 Objeto de Estudio.....	13
1.8 Ideas a Defender	13
1.9 Hipótesis	13
CAPÍTULO II. JUSTIFICACIÓN.....	14
CAPÍTULO III. FUNDAMENTACIÓN TEÓRICA	15
3.1 Desempeño Académico	15
3.2 Bases de Datos	15
3.3 Minería de Datos.....	16
3.4 Netbeans	26
3.5 AutoIt	26
CAPÍTULO IV. VALIDACIÓN EXPERIMENTAL.....	28
4.1 Comprensión del negocio	28
4.2 Comprensión de los datos	29
4.3 Preparación de los Datos.....	51
4.4 Modelado	56
4.5 Evaluación	64
4.6 Implementación	65
CAPÍTULO V. CONCLUSIÓN	67
REFERENCIAS	68



ÍNDICE DE FIGURAS

<i>Figura 1. Técnicas de minería de datos</i>	<i>18</i>
<i>Figura 2. Ejemplo de árbol de decisión.....</i>	<i>20</i>
<i>Figura 3. Fases del modelo de CRISP-DM.</i>	<i>22</i>
<i>Figura 4. WEKA</i>	<i>26</i>
<i>Figura 5. Atributos del CENEVAL y trayectorias ordenados por importancia</i>	<i>47</i>
<i>Figura 6. Estadística descriptiva PROM_BAC</i>	<i>48</i>
<i>Figura 7. Estadística descriptiva EDO_PROC</i>	<i>48</i>
<i>Figura 8. Estadística descriptiva CIU_PROC</i>	<i>49</i>
<i>Figura 9. Estadística descriptiva BEC_DAC</i>	<i>50</i>
<i>Figura 10. WEKA atributos e instancias</i>	<i>56</i>
<i>Figura 11. Selección del algoritmo de clasificación</i>	<i>57</i>
<i>Figura 12. Árbol de decisión 1</i>	<i>58</i>
<i>Figura 13. Árbol de decisión 2</i>	<i>59</i>
<i>Figura 14. Árbol de decisión 3</i>	<i>61</i>
<i>Figura 15. Árbol de decisión 4</i>	<i>62</i>
<i>Figura 16. Árbol de decisión 5</i>	<i>63</i>
<i>Figura 17. Árbol de decisión 6</i>	<i>63</i>
<i>Figura 18. Sistema de predicción</i>	<i>65</i>
<i>Figura 19. Ejemplo de predicción</i>	<i>65</i>
<i>Figura 20. Rangos del promedio de bachilleres.....</i>	<i>66</i>



ÍNDICE DE TABLAS

<i>Tabla 1. Descripción de los datos del CENEVAL</i>	<i>30</i>
<i>Tabla 2. Descripción de los datos de las trayectorias.....</i>	<i>40</i>
<i>Tabla 3. Fuentes de datos y sus instancias.....</i>	<i>50</i>
<i>Tabla 4. Descripción del tipo de atributo.....</i>	<i>53</i>
<i>Tabla 5. Descripción del atributo y su tipo</i>	<i>53</i>
<i>Tabla 6. Descripción de los datos</i>	<i>55</i>
<i>Tabla 7. Descripción de los datos</i>	<i>57</i>
<i>Tabla 8. Descripción de los datos</i>	<i>64</i>

CAPÍTULO I. INTRODUCCIÓN

Actualmente en la educación superior en América Latina muchos alumnos no pueden terminar sus estudios, algunas de las posibles causas son debido a situaciones familiares o sociales, falta de recursos, problemas con las materias, ya sea por falta de interés o dificultad que ésta le presente al estudiante, inclusive por problemas con algún docente, entre otras (Toledo, Romero, & Pereira, 2013). Esto es algo preocupante para las instituciones educativas ya que es de su interés que los alumnos concluyan sus estudios y en muchos casos es difícil el identificar a los estudiantes que puedan llegar a presentar un desempeño escolar bajo.

Estos problemas exigen soluciones. Para esto se han realizado diversos estudios, así como varias medidas para dar solución a este problema. Entre estas medidas se encuentran cambios estructurales en los sistemas educativos, asesorías a los alumnos y programas de recuperación, donde lo único que han generado son resultados poco claros y más costos para la institución educativa.

Como se expone anteriormente, las causas cubren un campo extensamente amplio, agregando el hecho de que cada alumno por separado puede presentar una combinación de estos factores dificultando el tomar una medida generalizada para combatir los problemas relacionados con los bajos niveles en el rendimiento académico.

Esto da pie a la presente investigación, en donde se busca predecir el desempeño académico de los alumnos en su primer cuatrimestre mediante el uso de minería de datos en la institución de la Universidad Tecnológica de Chihuahua (UTCH).

1.2 Antecedentes

Existe un creciente interés por determinar los múltiples factores que pueden influir en el desempeño académico. La mayoría de los trabajos que intentan resolver este problema están enfocados en determinar cuáles son los factores que más afectan al rendimiento de los estudiantes mediante la utilización de grandes cantidades de información generadas por los

alumnos. El problema es que identificar y encontrar información útil y oculta en grandes bases de datos es una tarea difícil. Una solución muy prometedora para alcanzar este objetivo es el uso de técnicas de extracción de conocimiento o minería de datos en la educación (Romero Morales, Márquez Vera, & Ventura Soto, 2012).

Utilizando técnicas de minería de datos, puede obtenerse un modelo del desempeño académico en la unidad académica deseada. Para identificar las características más relevantes se analiza, luego del pre-procesamiento de los datos, las proyecciones de los atributos en las clases o respuestas esperadas. Permitiendo así efectuar recomendaciones para reducir el porcentaje de alumnos con bajo desempeño académico (Formia, Lanzarini, & Hasperué, 2013).

Se realizaron investigaciones sobre los antecedentes relacionados al desempeño escolar en los que se manejan técnicas de minería de datos, indicando los problemas que se encontraron y las soluciones que se les dieron. A continuación se presenta lo encontrado.

En la Universidad de Simón Bolívar (Amaya, Barrientos, & Heredia, 2014) se realizó un modelo predictivo de deserción estudiantil, caracterizando a los estudiantes con el objetivo de poder predecir la probabilidad de su deserción. Se tuvieron en cuenta variables de tipo personal como edad actual, ciudad de procedencia, estado civil, nivel de estudios del padre, entre otras y de carácter académico como semestre, jornada, materias cursadas, materias perdidas y promedio. El modelo demostró el desempeño de los algoritmos presentados para clasificar datos. Para la creación del modelo se utilizó la herramienta de minería de datos WEKA y los árboles de decisión, ya que resultaron ser buenos clasificadores. Con el algoritmo J48 se produce un árbol de decisión de 6 reglas a partir de 201 instancias las cuales fueron utilizadas como conjunto de entrenamiento y para probar el modelo obtenido. Como resultado de predicción se observó que el modelo mostró

a 21 estudiantes en riesgo de reprobación el segundo semestre del año 2011, de esos 21 en realidad fueron 9 los acertados por el sistema de predicción.

En la Universidad de IUCESMAG y Nariño (Toledo, Romero, & Pereira, 2013), se detectaron patrones de bajo desempeño académico utilizando técnicas de minería de datos y tomando como referencia datos socioeconómicos, académicos, disciplinarios e institucionales. Se construyó un repositorio de datos con la información de los estudiantes que ingresaron en el primer y segundo semestre en los años 2004 y 2006, además se utilizaron técnicas de clasificación y *clustering* con la herramienta WEKA sobre los datos de los estudiantes obteniendo como resultado un patrón común determinado por un promedio bajo y la pérdida de materias en los primeros semestres de la carrera. Al construir un árbol de decisión sobre el conjunto de datos de los alumnos, se obtuvieron reglas de clasificación generales con una confianza mayor de 75% para los rangos de notas entre 0 y 5 indicando que si el promedio de notas es menor que 2.4 el estudiante tiene un desempeño académico bajo y abandona sus estudios. El conocimiento generado permite soportar la toma de decisiones eficaces por parte de las directivas universitarias enfocadas a formular políticas y estrategias relacionadas con los programas de retención de estudiantil.

En la universidad Francisco de Paula Santander Ocaña, Colombia (Gómez, 2013), se utiliza la minería de datos donde se determinaron factores que predicen un desempeño académico bajo, los cuales fueron la edad, promedio, ciudad, la fecha de ingreso y el semestre que cursa. Se utilizaron técnicas de clasificación con la herramienta WEKA donde se evaluó el conjunto de datos agrupando todos los programas académicos, dando una precisión del 90,6% de las instancias. Dentro del modelo obtenido se obtuvieron las siguientes reglas que determinaron si el estudiante obtendría un desempeño académico bajo:

1. Si el promedio general es aceptable el cual se encuentra dentro del rango de 3 y 3.5 en una escala del 0 al 5, la ciudad de origen es Aguachica, la edad es mayor de 26 y está en los primeros semestres, entonces el estudiante tendrá un desempeño académico bajo.
2. Si el promedio general es aceptable, la ciudad de origen es Aguachica, el programa es Ingeniería Mecánica y el promedio del periodo es aceptable, entonces el estudiante tendrá un desempeño académico bajo.
3. Si el promedio general es aceptable, es de los primeros semestres, la ciudad de origen es Abrego, la fecha de ingreso fue entre el 2005 y 2008, entonces el estudiante puede tener un desempeño académico bajo.
4. Si el promedio general es aceptable, los créditos aprobados son pocos y la fecha de ingreso está entre el 2000 y 2005, entonces el estudiante tendrá un desempeño académico bajo.

Las instituciones de nivel superior en México comúnmente utilizan los Exámenes Nacionales de Ingreso (EXANI), para identificar a los alumnos más capacitados que deseen entrar en sus instituciones. Esto le genera confianza a la institución de que un alto porcentaje de los alumnos con los mejores resultados en el EXANI no tendrán un desempeño bajo y que podrán terminar sus estudios (CENEVAL, 2015).

Esta investigación utiliza como caso de estudio a alumnos de la Universidad Tecnológica de Chihuahua (UTCH). Se analiza la información que la institución nos brinda de los estudiantes que cursaron el primer cuatrimestre. Se usa la herramienta WEKA y varios modelos computacionales con el fin de predecir si un alumno tiene alta probabilidad de tener un rendimiento bajo, y así poder tomar medidas preventivas para que siga con sus estudios.

1.3 Problema de Investigación

En la UTCH se tiene el problema de que es difícil el identificar si un alumno tendrá un desempeño académico bajo. Al tratar de resolver estos problemas generan más costos para la institución ya que se dan asesorías a alumnos que nos las necesitan y se le tiene que pagar al maestro que las imparte.

1.4 Pregunta de Investigación

¿Cómo predecir el desempeño académico de los alumnos en el primer cuatrimestre de la UTCH?

1.5 Objetivos

En este punto se explica el objetivo general del proyecto, así como los objetivos específicos, con el fin de exponer lo que pretende lograrse ya que todo el trabajo realizado se enfoca a lograr estos objetivos.

1.5.1 Objetivo general

Desarrollar una herramienta de estimación del desempeño académico a partir de la generación de un modelo predictivo de clasificación.

1.5.2 Objetivos específicos

1. Utilizar la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*) para el desarrollo del proyecto.
2. Búsqueda de los modelos de clasificación más precisos y comprensibles usando el principio MDL (*Minimum Description Length*), para estimar el desempeño académico de los alumnos de nuevo ingreso.
3. Validar la utilidad y novedad de los modelos encontrados, definida por el juicio expertos.

4. Desarrollar un sistema de predicción por clasificación para auxiliar en la toma de decisiones sobre el desempeño académico de los alumnos en el primer cuatrimestre.

1.6 Aportación Práctica

Al predecir que alumnos tendrán desempeño académico bajo se les puede brindar ayuda personalizada para que puedan mejorar su desempeño académico y así lograr la conclusión de sus estudios.

1.7 Objeto de Estudio

Como objeto de estudio son los alumnos de primer cuatrimestre de la UTCH en el periodo del 2013 al 2015. En específico sus datos históricos y los resultados de los exámenes de ingreso EXANI aplicados por el CENEVAL.

1.8 Ideas a Defender

- La selección de ingreso de los alumnos se puede mejorar si se toman en cuenta los atributos más importantes que definen su desempeño académico.
- Los sistemas de predicción por clasificación son buenas herramientas que permiten obtener modelos precisos, comprensibles, útiles y novedosos.

1.9 Hipótesis

La hipótesis a comprobar es si se puede estimar la predicción del desempeño académico de los alumnos en su primer cuatrimestre de la UTCH mediante un sistema de predicción por clasificación de alta calidad. La alta calidad es medida por una precisión del modelo superior o igual a 85%, el modelo es comprensible con no más de 5 reglas, el modelo es útil para determinar el bajo desempeño académico de los alumnos de primer cuatrimestre y las reglas generadas son novedosas a consideración de los expertos.



CAPÍTULO II. JUSTIFICACIÓN

Las instituciones de nivel superior tienen como objetivo principal mejorar el desempeño académico de sus estudiantes además de ayudarlos a que puedan terminar con sus estudios. Para esto es necesario anticiparse a los problemas de los alumnos que se les lleguen a presentar.

Por lo cual se creará un sistema de predicción usando técnicas y modelos de clasificación, para estimar la probabilidad de que un alumno tenga un desempeño académico alto, medio, o bajo. Se utiliza como caso de estudio a los alumnos del primer cuatrimestre de la institución de la UTCH.

Al llevar a cabo éste proyecto de investigación la UTCH contará con la ayuda de un sistema de predicción para identificar a los alumnos con bajo desempeño académico y así poder tomar las medidas necesarias para que continúen con sus estudios. También auxilia en la mejora de la calidad del sistema educativo ya que pueden tomar decisiones eficaces por parte de la institución al elaborar nuevas estrategias y métodos de enseñanza basados en la información generada por los alumnos que cursen carreras en dicha institución.

CAPÍTULO III. FUNDAMENTACIÓN TEÓRICA

Para tener una mayor comprensión del funcionamiento y los requerimientos del uso de la minería de datos en el desempeño estudiantil, es necesario conocer sus conceptos fundamentales. En esta sección se explican tales conceptos, así como la terminología más utilizada en este ámbito para el entendimiento de los temas que se abordarán en esta investigación.

3.1 Desempeño Académico

A continuación se presentan diferentes definiciones del concepto de desempeño académico:

- (Figuerola, 2004): Define el desempeño académico como el producto de la asimilación del contenido de los programas de estudio, expresado en calificaciones dentro de una escala convencional.
- (Montes & Lerner, 2010): Definen el desempeño académico como la relación entre el proceso de aprendizaje, que involucra factores extrínsecos e intrínsecos al individuo, y el producto que se deriva de él, expresado tanto en valores predeterminados por un contexto sociocultural como en las decisiones y acciones del sujeto en relación con el conocimiento que se espera obtenga de dicho proceso”.

En esta investigación el desempeño o rendimiento académico se define como el resultado cuantitativo (expresado en una escala discreta) que se obtiene en el proceso de aprendizaje de conocimientos en el ámbito escolar, conforme a las evaluaciones que realiza el docente mediante pruebas objetivas y otras actividades complementarias.

3.2 Bases de Datos

Las bases de datos se pueden definir como:

- Una colección de datos interrelacionados entre sí que se encuentran almacenados para facilitar su acceso de manera más eficiente por medio de programas.
- Un conjunto de datos organizados con características similares, que muestran una característica común.
- Una colección de información que se encuentre interrelacionada y muestre datos descriptivos.

En nuestra investigación una base de datos es el conjunto de datos organizados dentro de un archivo con características similares en este caso los datos de los alumnos de la UTCH.

3.2.1 Componentes de un sistema de base de datos

Los componentes de un sistema de base de datos son (Greiner, 2014):

- Los datos: Son los elementos individuales de los archivos, también son denominados campos.
- Registros: Un registro es el conjunto completo de datos relacionados pertenecientes a una entrada. Estos registros tienen fijos su longitud y tipos (alfanuméricos, numéricos, o alfabéticos).
- Llave del registro: Se utiliza para distinguir un registro de otro, al elegir uno de los datos cuyo valor sea único en todos los registros del archivo.
- Entidad: Una entidad es cualquier persona, lugar, cosa o evento de interés para la organización y acerca del cual se capturan, almacenan o procesan datos.
- Archivo: Un archivo es una colección de registros relacionados. Se incluye cada registro en un archivo ya que pertenece a la misma entidad.

3.3 Minería de Datos

La mayor razón por la que la minería de datos ha atraído la atención en la industria de la información en los últimos años se debe a la amplia disponibilidad de grandes

cantidades de datos en las empresas y a la inminente necesidad de convertir dichos datos en información y conocimiento útiles.

A continuación se define la minería de datos y sus principales características. Así mismo, se describe la metodología CRISP-DM en la cual se basará la investigación para la realización del proyecto de minería de datos y la herramienta utilizada de minería de datos.

3.3.1 Definición

(García, 2012) Define la minería de datos a la extracción de conocimiento a partir de grandes cantidades de datos, sobre los que se aplican métodos inteligentes con el fin de extraer patrones interesantes. Dichos datos pueden estar almacenados en bases de datos, almacenes de datos u otros repositorios de información.

En esta investigación la minería de datos se define como el proceso de descubrir conocimiento relevante, por medio de asociaciones, patrones, anomalías, cambios y estructuras analizando la información que se encuentre almacenada en bases de datos, datawarehouses, o de cualquier forma en que se encuentre guardada.

3.3.2 Técnicas de minería de datos

La minería de datos tiene como objetivo analizar los datos para extraer conocimiento. Este conocimiento puede ser en forma de relaciones, patrones o reglas inferidos de los datos (previamente) desconocidos, o bien en forma de una descripción más concisa. Estas relaciones o resúmenes constituyen el modelo de los datos analizados. Existen muchas formas diferentes de representar los modelos y cada una de ellas determina el tipo de técnica que puede usarse para inferirlos (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004).

- Las técnicas predictivas pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variable objetivo o dependientes, usando otras variables o campos de la base de datos, a los que nos referimos como variables independientes o predictivas.

- Las técnicas descriptivas identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos.
- Las técnicas auxiliares que son herramientas de apoyo más superficiales y limitadas.

Las técnicas de minería de datos se clasifican en dos grandes categorías: supervisadas o predictivas y no supervisadas o descriptivas. En la siguiente figura se ilustran los tipos de técnicas existentes clasificadas por categoría (ver Figura 1. Técnicas de minería de datos).

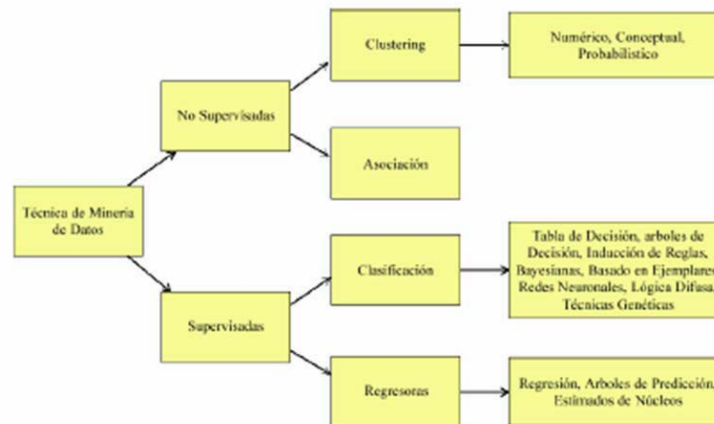


Figura 1. Técnicas de minería de datos

Una técnica constituye el enfoque conceptual para extraer la información de los datos, y, en general es implementada por varios algoritmos. Cada algoritmo representa, en la práctica, la manera de desarrollar una determinada técnica paso a paso.

Para el desarrollo del análisis de los datos proporcionados por la UTCH, se utilizará una técnica de predicción por clasificación para identificar a los alumnos con bajo desempeño escolar. La fórmula para el cálculo de la precisión que se utiliza es:

$$\text{precisión} = \frac{\text{verdaderos positivos}}{\text{No. de positivos predecidos}}$$

3.3.3 *NaiveBayes*

NaiveBayes pertenece a la clase de clasificadores estadísticos que predicen la clase de pertenencia de un ejemplo X mediante probabilidades, la probabilidad que un ejemplo dado pertenezca a una clase particular, se basa en el supuesto que los efectos de un atributo son independientes de los valores de los otros atributos de un ejemplo X . Para problemas de clasificación se desea determinar $P(H|X)$, la probabilidad de que la hipótesis sostenida da al ejemplo X . El teorema de Bayes proporciona un algoritmo eficiente para calcular esta probabilidad a priori $P(H|X)$. Comparado con las redes bayesianas el clasificador es mucho más rápido y si los atributos son independientes puede tener la mayor tasa de acierto comparado con otros clasificadores (Suca, Córdova, Condori, Cayra, & Sulla, 2016).

3.3.4 *LibSVM*

Es una implementación para la máquina de vectores soporte (SVM), cuyo objetivo es aplicar fácilmente el clasificador SVM en las aplicaciones, teniendo una gran popularidad en cuanto al aprendizaje automático. El uso típico de LibSVM en su tarea de clasificador contiene dos pasos: primero, entrenamiento de un conjunto de datos para obtener un modelo y segundo, usar el modelo obtenido para predecir información de un conjunto de datos, tanto para SVC(clasificación) y SVR(Regresión), LibSVM aparte de entregar la precisión del sistema, puede también tener como salida: probability estimates, que puede interpretarse como la probabilidad estimada de acierto(verdadero positivo) o de error (falso positivo) (Paute, 2016).

El objetivo de las SVM es producir un modelo basado en los datos de entrenamiento que prediga las clases o categorías de un conjunto nuevo de instancias, mediante la generación de un hiperplano en un espacio de n -dimensión. Las SVM funcionan bien para clasificar texto, porque cuando se clasifica texto se trabaja con espacios de dimensión alta, con pocas características irrelevantes, con representaciones vectoriales dispersas y porque la mayor parte de problemas de clasificación de texto son linealmente separables. Todos

estos argumentos son fundamentos teóricos que nos permiten saber que es posible clasificar texto con esta clase de algoritmos supervisados (Garibay & Galicia, 2015).

3.3.5 Árboles de Decisión

Los árboles de decisión se conocen como un algoritmo de aprendizaje basado en similitudes, además de ser uno de los algoritmos más sencillos y fáciles de implementar pero que son de los que más brindan ayuda, ya que generan un árbol de decisión en forma recursiva y consideran la mayor cantidad de los datos que sean más precisos para obtener la mejor información, al elegir el atributo que mejor clasifica a los datos (Valero, García, & Salvador, 2010).

Un árbol de decisión indica las acciones a realizar en función del valor de una o varias variables. Es una representación en forma de árbol cuyas ramas se bifurcan en función de los valores tomados por las variables y que terminan en una acción concreta. Se suele utilizar cuando el número de condiciones no es muy grande (en tal caso, es mejor utilizar una tabla de decisión).

La siguiente figura muestra un ejemplo de un árbol de decisión (ver Figura 2. Ejemplo de árbol de decisión).

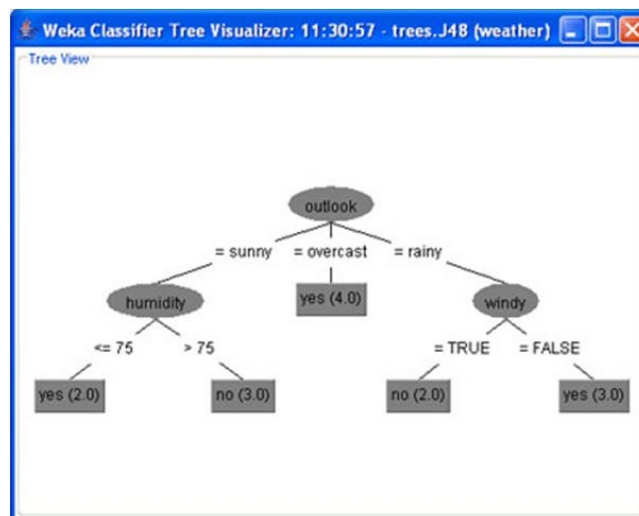


Figura 2. Ejemplo de árbol de decisión.

3.3.5.1 Sistemas de reglas en los árboles de decisión

El aprendizaje inductivo es un caso particular entre las técnicas de aprendizaje a partir de ejemplos, siendo su cometido el inducir reglas a partir de los datos históricos disponibles.

La inducción del árbol de decisión está estrechamente relacionada con la regla de inducción. Cada camino desde la raíz de un árbol de decisión a una de sus hojas se puede transformar en una regla simplemente mediante la conjunción de las pruebas a lo largo de la ruta para formar la parte antecedente y tomando la clase de predicción de la hoja como el valor de la clase (Rokach & Maimon, 2015).

Para la aplicación de estos sistemas de aprendizaje inductivo se parte de un conjunto de ejemplos. Cada ejemplo debe tener la misma estructura consistente en una conclusión (o decisión) y un número de características o atributos que definen esa conclusión o decisión.

El sistema construye un árbol de decisión que representa la relación existente entre la conclusión-decisión y sus atributos. Es decir, se produce un proceso de generalización de forma que el árbol de decisión generado clasifica correctamente los ejemplos dados. Este árbol, además, se caracteriza por ser el óptimo en el sentido que minimiza el número de atributos requeridos para alcanzar la conclusión-decisión (Bonsón, Escobar, & Zamora).

3.3.6 Metodología y guía de referencia CRISP-DM

CRISP-DM es un método probado para orientar los trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos.

El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario. El modelo de CRISP-DM es flexible y se puede personalizar fácilmente (ver Figura 3. Fases del modelo de CRISP-DM.).

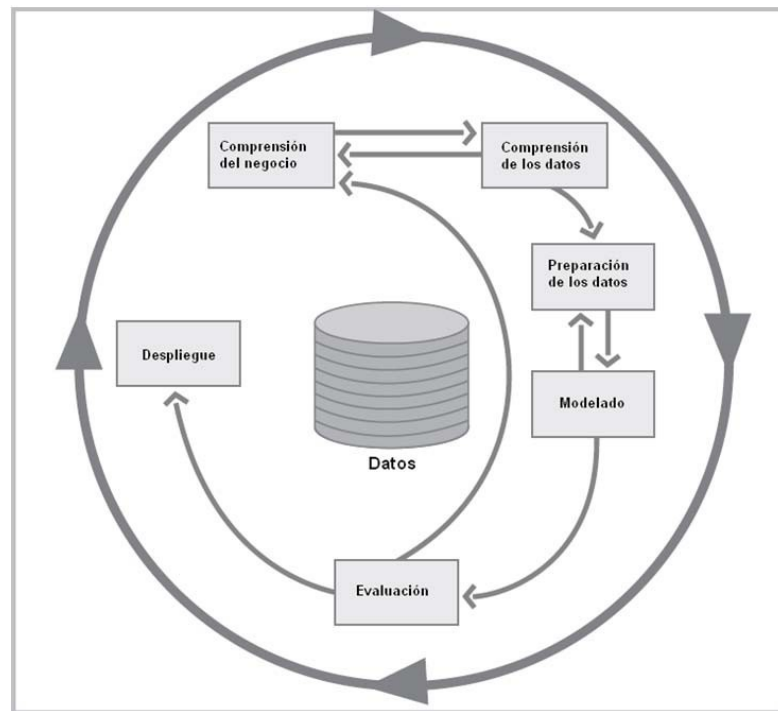


Figura 3. Fases del modelo de CRISP-DM.

3.3.6.1 Comprensión del negocio

La primera fase se centra en entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio, luego convirtiendo este conocimiento de los datos en la definición de un problema de minería de datos y posteriormente en un plan preliminar diseñado para alcanzar los objetivos. Esta fase consta de cuatro sub-fases:

1. Establecimiento de los objetivos de negocio.
2. Evaluación de la situación.
3. Establecimiento de los objetivos de minería de datos.
4. Generación del plan del proyecto.

3.3.6.2 Comprensión de los datos

Esta etapa busca recopilar y familiarizarse con los datos del negocio, identificar los problemas de calidad de datos, identificar los problemas de calidad que pudieran tener los datos, descubrir los primeros conocimientos en los datos, y descubrir posibles subconjuntos

de datos para poder formular hipótesis en cuanto a la información oculta. La fase de comprensión de los datos consta de cuatro sub-fases:

1. Recopilación inicial de datos.
2. Descripción de datos.
3. Exploración de datos.
4. Verificación de calidad de datos.

3.3.6.3 Preparación de los Datos

Esta fase cubre todas las actividades necesarias para construir el conjunto de datos final. Las tareas de preparación de datos probablemente van a ser realizadas muchas veces y no en cualquier orden. Aquí se incluye la integración, selección, limpieza y transformación. Consta de cinco sub-fases:

1. Selección de datos.
2. Limpieza de datos.
3. Construcción de datos.
4. Integración de datos.

3.3.6.4 Modelado

Es la aplicación de varias técnicas de modelado o de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos; por lo tanto volver a la fase de preparación de los datos es a menudo necesario. Consta de cuatro sub-fases:

1. Selección de la técnica de modelado.
2. Diseño de la evaluación.
3. Construcción del modelo.
4. Evaluación del modelo.

3.3.6.5 Evaluación

Es necesario evaluar los modelos obtenidos de la fase anterior, pero ya no solo desde un punto de vista estadístico respecto a los datos, como se realiza en la última sub-fase de la fase anterior, sino ver si el modelo se ajusta a las necesidades establecidas en la

primera fase, es decir, si el modelo nos sirve para responder a algunos de los requerimientos del negocio. Antes de proceder con la evaluación del modelo que se ha creado, es importante evaluar a fondo la ejecución de los pasos para crearlo, para comparar el modelo correctamente obtenido con los objetivos de negocio. Consta de tres sub-fases:

1. Evaluación de resultados.
2. Revisar el proceso.
3. Establecimiento de los siguientes pasos.

3.3.6.6 Despliegue o Implementación

Esta etapa tiene como objetivo el implementar una aplicación que tenga una interfaz intuitiva y fácil de usar por los docentes o usuarios finales encargados de la gestión estudiantil. Esta aplicación hará uso de la información recabada y procesada, con el objetivo de que se pueda detectar de manera más fácil si un alumno tiene probabilidades de desertar.

Esta fase también se le conoce como desarrollo y principalmente se trata de explotar la potencialidad de los modelos, integrarlos en los procesos de toma de decisión de la organización, difundir informes sobre el conocimiento extraído, etc. la creación del modelo no es generalmente el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento ganado tendrá que ser organizado y presentado en el modo en el que el cliente pueda usarlo. Consta de cuatro sub-fases:

1. Planificación del despliegue.
2. Planificación de la monitorización y del mantenimiento.
3. Generación del informe final.
4. Revisión del proyecto.

Lo destacable del estándar CRISP-DM es, en primer lugar, el detalle de las fases y la especificación de los subproductos de cada sub-fase. El desarrollo de cada fase conlleva aspectos más específicos que se describen en cada una de sus sub-fases correspondientes.

La especificación del CRISP-DM, en especial la guía, detalla cada una de estas fases; se puede obtener una copia de esta guía y adaptarla a las necesidades de cualquier empresa, siempre y cuando el estudio e implantación sea lo más adecuado posible para lograr su éxito.

3.3.7 WEKA

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

El paquete WEKA contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades (Correa Quezada, 2010).

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



Figura 4. WEKA

3.4 Netbeans

Netbeans es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Existe además un número importante de módulos para extenderlo. Siendo un proyecto de código abierto. Sun Microsystems fundó el proyecto de código abierto Netbeans en Junio de 2000 y continúa siendo el patrocinador principal de los proyectos.

La plataforma Netbeans permite que las aplicaciones sean desarrolladas a partir de un conjunto de componentes de software llamados módulos. Un módulo es un archivo Java que contiene clases de Java escritas para interactuar con las APIs de Netbeans y un archivo especial que lo identifica como módulo. Las aplicaciones construidas a partir de módulos pueden ser extendidas agregándole nuevos módulos. Debido a que los módulos pueden ser desarrollados independientemente, las aplicaciones basadas en la plataforma Netbeans pueden ser extendidas fácilmente por otros desarrolladores de software.

3.5 AutoIt

AutoIt es un lenguaje de programación similar a BASIC gratuito diseñado para automatizar la interfaz gráfica de usuario de Windows y scripting en general. Utiliza una combinación de pulsaciones de teclas simuladas, movimiento del ratón y de la ventana así como el control de manipulación con el fin de automatizar las tareas de una manera que no es posible o fiable con otros lenguajes de programación. AutoIt es también muy pequeño, autónomo y se ejecutará en todas las versiones de Windows sin los molestos "tiempos de ejecución" (AutoIt, 2016).

AutoIt maneja las siguientes características:

- Fácil de aprender la sintaxis similar al BASIC.
- Simula las pulsaciones del teclado y movimientos del ratón.



- Manipular ventanas y procesos.
- Interactuar con todos los controles estándar de Windows.
- Los scripts pueden ser compilados en ejecutables independientes.
- Crea interfaces gráficas de usuario.
- Soporta COM.
- Expresiones regulares.
- Llama directamente a DLL externa y las funciones de la API de Windows.
- Archivo de ayuda detallado y grandes foros de soporte basados en la comunidad
- El soporte Unicode y x64
- Firmado digitalmente
- Funciona con el Control de cuentas de usuario de Windows Vista (UAC)

CAPÍTULO IV. VALIDACIÓN EXPERIMENTAL

En este capítulo se muestra el modelo de clasificación de calidad obtenido.

Se realizó un proceso de limpieza y preparación de los datos de origen, al utilizar la información histórica del CENEVAL así como de las trayectorias realizadas a los alumnos de la UTCH.

La realización del proyecto de minería de datos se efectúa mediante el seguimiento de las fases sugeridas por la metodología CRISP-DM, las cuales se detallan a continuación.

4.1 Comprensión del negocio

Lo primero que nos interesa es definir el problema. A continuación se muestra el problema desde el punto de vista respecto a la empresa y respecto a la minería de datos.

4.1.1 *Objetivos del negocio*

El objetivo del negocio es mejorar la formación y el desempeño escolar de los alumnos matriculados en la institución, además de ayudarlos a que puedan terminar con sus estudios. Para que al término de sus estudios puedan ser competentes en el ámbito laboral de la profesión ejercida.

Desde el punto de vista de la minería de datos se busca predecir a los alumnos con bajo desempeño escolar haciendo uso de datos históricos.

4.1.1.1 *Valoración de la situación*

Actualmente la UTCH cuenta con la información histórica del CENEVAL, registros de las calificaciones de los alumnos y trayectorias de información recabada de los alumnos que cursaron el primer cuatrimestre. Esta información es necesaria para el análisis de los datos pero no se encuentra ordenada; esto quiere decir que la información de los alumnos no se encuentra unificada ni capturada en un formato o sistema administrable.

4.1.1.2 *Área del problema*

La UTCH difícilmente puede identificar a los alumnos que tendrán un desempeño académico bajo además el tratar de resolver este problema genera más gastos económicos

para la institución y muchas veces no se logra el resultado esperado. Esto tiene como consecuencia que muchos alumnos obtienen un desempeño académico bajo.

4.1.1.3 Objetivos de la minería de datos

Los objetivos en términos de minería de datos son:

1. Análisis y pre-procesamiento de los datos.
2. Determinar los atributos que tengan mayor importancia para la predicción.
3. Utilizar las técnicas de clasificación para encontrar el modelo de mayor calidad para tomarlo como base para predecir el desempeño escolar.

4.1.1.4 Producción del plan del proyecto

Para el plan del proyecto primeramente se realiza un análisis y el pre-procesamiento de los datos proporcionados por la UTCH. Una vez realizado se crearan repositorios de los datos para ser analizada en WEKA y así obtener los modelos necesarios para la predicción, con validación por expertos de la UCTH para los parámetros de calidad subjetivo: utilidad y novedad. Por último se creara un programa en el cual se use el modelo elegido por los expertos y este servirá para realizar las predicciones de los estudiantes de nuevo ingreso en la institución.

4.2 Comprensión de los datos

Para poder interpretar los resultados de la investigación es necesario recordar que el proceso de experimentación se abordó uniendo la información recabada de dos partes, una de ellas contiene toda la información referente a los exámenes de admisión CENEVAL y la otra parte contiene la información que la UTCH haya recabado de los alumnos del primer cuatrimestre.

4.2.1 Recolección de datos

Los datos relacionados al primer cuatrimestre se obtienen directamente de las bases de datos proporcionados por la institución, estos datos son:

- Promedio final del primer cuatrimestre.

- Examen CENEVAL.
- Trayectorias de los alumnos.

4.2.2 Transformación de los datos

Se crearan 2 repositorios de los datos obtenidos, estos estarán en formato CSV para poder analizarlos con WEKA y así poder crear un sólo repositorio con formato ARFF.

Para trabajar con WEKA es necesario transformar la información para que se puedan tomar los tipos de datos de los atributos ya sean nominales o numéricos, además de transformar el archivo a CSV ya que si se cargan en formato XLSX o XLS no serán reconocidos por el programa.

4.2.3 Descripción de los datos

A continuación se muestran las tablas de los datos recabados de la UTCH con los que se realiza la minería de datos:

Tabla 1. Descripción de los datos del CENEVAL

VARIABLE	DESCRIPCIÓN	VALORES	ETIQUETAS
I. IDENTIFICACIÓN DE LA APLICACIÓN			
TIPO_EXA	TIPO DE EXAMEN	EX2	
APLI	NÚMERO DE APLICACIÓN	000000000-999999999	
FECHA_APLI	FECHA DE APLICACIÓN		
II. IDENTIFICACIÓN DE LA INSTITUCIÓN SEDE			
CVE_INST	CLAVE DE LA INSTITUCIÓN EN DONDE SE LLEVÓ A CABO LA APLICACIÓN		
IDENTIFICA	IDENTIFICACIÓN DE LA APLICACIÓN	1-2: Entidad Federativa (01-33)	
		3-5: Plantel / Facultad (000-999)	
		6-8: Turno / carrera (000-999)	
III. IDENTIFICACIÓN DEL SUSTENTANTE			
FOLIO	FOLIO	1-7: Número consecutivo	

		8: Año de aplicación (3)	
		9: Dígito verificador (0-9)	
MATRICULA	ANOTE LA CLAVE O MATRÍCULA DE LA INSTITUCIÓN	1-10: Matrícula C.C. (0-9)	
		11-12: Adicionales (0-9,A-Z)	
IV. CUESTIONARIO DE CONTEXTO			
Sección 1. Datos generales			
APE_PAT	APELLIDO PATERNO	ABIERTA	
APE_MAT	APELLIDO MATERNO	ABIERTA	
NOMBRE	NOMBRE	ABIERTA	
FECHA DE NACIMIENTO			
DIA_NAC	DÍA DE NACIMIENTO	00-39	
MES_NAC	MES DE NACIMIENTO	01	Enero
		02	Febrero
		03	Marzo
		04	Abril
		05	Mayo
		06	Junio
		07	Julio
		08	Agosto
		09	Septiembre
		10	Octubre
		11	Noviembre
		12	Diciembre
ANO_NAC	AÑO DE NACIMIENTO	30-99	
SEXO	SEXO	1	Hombre
		2	Mujer
¿ALGUNO DE SUS PADRES HABLA UNA LENGUA INDÍGENA O DIALECTO?			
LI_MAD	MADRE	1	Sí
		2	No
		3	No sé
LI_PAD	PADRE	1	Sí
		2	No
		3	No sé
Sección 2. Datos escolares			
EDO_PROC	¿EN QUÉ ENTIDAD CONCLUYÓ EL NIVEL MEDIO SUPERIOR?	01	Aguascalientes
		02	Baja California
		03	Baja California Sur
		04	Campeche
		05	Coahuila
		06	Colima
		07	Chiapas
		08	Chihuahua
		09	Distrito Federal

		10	Durango
		11	Guanajuato
		12	Guerrero
		13	Hidalgo
		14	Jalisco
		15	México (Edo. de)
		16	Michoacán
		17	Morelos
		18	Nayarit
		19	Nuevo León
		20	Oaxaca
		21	Puebla
		22	Querétaro
		23	Quintana Roo
		24	San Luis Potosí
		25	Sinaloa
		26	Sonora
		27	Tabasco
		28	Tamaulipas
		29	Tlaxcala
		30	Veracruz
		31	Yucatán
		32	Zacatecas
		99	Extranjero
IDENTIFICACIÓN DE LA INSTITUCIÓN DONDE CONCLUYÓ EL NIVEL MEDIO SUPERIOR			
NOM_PROC	NOMBRE DE LA INSTITUCIÓN EN LA QUE CONCLUYÓ EL NIVEL MEDIO SUPERIOR	ABIERTA	
CIU_PROC	NOMBRE DE LA CIUDAD DONDE SE UBICA LA INSTITUCIÓN EN LA QUE CONCLUYÓ EL NIVEL MEDIO SUPERIOR; SI FUE EN EL EXTRANJERO INDIQUE EL PAÍS.	ABIERTA	
CVE_PROC	CLAVE DE LA INSTITUCIÓN EN LA QUE CONCLUYÓ SUS ESTUDIOS DE BACHILLERATO	1-5 (consecutivo)	
		6 (dígito verificador)	
ANO_BAC	¿EN QUÉ AÑO CONCLUYÓ EL NIVEL MEDIO SUPERIOR?	1	2011 o antes
		2	2012
		3	2013
REG_PROC	¿CUÁL ES EL RÉGIMEN DE LA ESCUELA EN LA QUE ESTUDIÓ EL NIVEL MEDIO SUPERIOR?	1	Pública
		2	Privada
		3	Federal por cooperación
MOD_BAC	¿EN QUÉ MODALIDAD OBTUVO SU CERTIFICADO DEL BACHILLERATO?	1	Bachillerato general
		2	Bachillerato tecnológico
		3	Profesional técnico bachiller
		4	Bachillerato intercultural (bilingüe indígena)
		5	Bachillerato

			internacional
		6	Acuerdo secretarial 286 Bachillerato
		7	Preparatoria abierta
		8	Educación media superior a distancia
PROM_BAC	¿CUÁL FUE SU PROMEDIO GENERAL EN EL NIVEL MEDIO SUPERIOR?	1	6.0 - 6.4
		2	6.5 - 6.9
		3	7.0 - 7.4
		4	7.5 - 7.9
		5	8.0 - 8.4
		6	8.5 - 8.9
		7	9.0 - 9.4
		8	9.5 - 9.9
		9	10
¿RECIBIÓ BECA EN EL NIVEL MEDIO SUPERIOR?			
BEC_DAC	POR DESEMPEÑO ACADÉMICO	1	Sí
		2	No
BEC_NEC	POR NECESIDAD ECONÓMICA	1	Sí
		2	No
BEC_HDA	POR HABILIDAD DEPORTIVA O ARTÍSTICA	1	Sí
		2	No
Sección 3. Situación Laboral			
HRS_TRAB	MIENTRAS ESTUDIABA EL BACHILLERATO, ¿CUÁNTAS HORAS A LA SEMANA DEDICABA A TRABAJAR?	1	No trabajaba
		2	Menos de 5 horas
		3	De 5 a 10 horas
		4	De 11 a 15 horas
		5	De 16 a 20 horas
		6	Más de 20 horas
Sección 4. Características Personales			
EST_ALCA	¿CUÁL ES EL NIVEL MÁXIMO DE ESTUDIOS QUE LE GUSTARÍA ALCANZAR?	1	Carrera de técnico superior universitario
		2	Licenciatura
		3	Posgrado (especialidad, maestría, doctorado)
NO_UNI	SI NO SE GRADUARA DE UNA CARRERA UNIVERSITARIA, ¿QUÉ INGRESO MENSUAL PROMEDIO ESPERARÍA OBTENER DENTRO DE 10 AÑOS?	1	\$4,000 o menos
		2	\$4,001 a \$7,000
		3	\$7,001 a \$10,000
		4	\$10,001 a \$15,000
		5	\$15,001 a \$20,000
		6	\$20,001 a \$30,000
		7	Más de \$30,000
SI_UNI	SI SE GRADUARA DE UNA CARRERA UNIVERSITARIA, ¿QUÉ INGRESO MENSUAL PROMEDIO ESPERARÍA OBTENER DENTRO DE 10 AÑOS?	1	\$4,000 o menos
		2	\$4,001 a \$7,000
		3	\$7,001 a \$10,000
		4	\$10,001 a \$15,000
		5	\$15,001 a \$20,000
		6	\$20,001 a \$30,000
		7	Más de \$30,000
SI_POSG	SI SE GRADUARA DE UN POSGRADO, ¿QUÉ INGRESO MENSUAL PROMEDIO ESPERARÍA OBTENER DENTRO DE 10 AÑOS?	1	\$4,000 o menos
		2	\$4,001 a \$7,000
		3	\$7,001 a \$10,000

		4	\$10,001 a \$15,000
		5	\$15,001 a \$20,000
		6	\$20,001 a \$30,000
		7	Más de \$30,000
INDIQUE QUÉ TANTO LO DESCRIBE LO SIGUIENTE:			
DES_IDEA	ALGUNA VEZ ME OBSESIONÉ CON UNA IDEA O META POR UN TIEMPO CORTO, PERO MÁS TARDE PERDÍ EL INTERÉS	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_DESA	LOS PROBLEMAS INESPERADOS NO ME DESANIMAN	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_ATEN	TENGO DIFICULTAD PARA MANTENER MI ATENCIÓN EN METAS QUE REQUIEREN VARIOS MESES PARA SER TERMINADAS	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_ESME	SOY UNA PERSONA QUE SE ESMERA	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_META	A MENUDO ME PASA QUE DESPUÉS DE TENER UNA META CAMBIO A OTRA	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_DIST	A VECES NUEVAS IDEAS O METAS ME DISTRAEN DE LAS QUE YA TENÍA ESTABLECIDAS	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_TERM	TERMINO TODO LO QUE EMPIEZO	1	No me describe
		2	Me describe poco
		3	Me describe
		4	Me describe mucho
		5	Me describe totalmente
DES_DURO	SOY UNA PERSONA QUE TRABAJA DURO	1	No me describe
		2	Me describe poco
		3	Me describe

		4	Me describe mucho
		5	Me describe totalmente
INDICA LA FRECUENCIA CON LA QUE REALIZAS LAS SIGUIENTES ACTIVIDADES:			
FRE_LIST	ELABORO UNA LISTA DE LAS ACTIVIDADES IMPORTANTES QUE TENGO QUE HACER	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_PLAN	PLANEO LAS ACTIVIDADES QUE HARÉ EN EL DÍA	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_ICLA	TENGO UNA IDEA CLARA DE LO QUE QUIERO LOGRAR DURANTE LA SIGUIENTE SEMANA	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_EPRIO	ESTABLEZCO MIS PRIORIDADES	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_CPRI	CUMPLO MIS PRIORIDADES	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_INTER	HAGO COSAS QUE INTERFIEREN CON MIS PRIORIDADES	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
Sección 5. Habilidad para escribir			
HAB_EIME	CALIFIQUE SU NIVEL DE HABILIDAD PARA EXPRESAR SUS IDEAS DE MANERA ESCRITA, EN COMPARACIÓN CON SUS COMPAÑEROS.	1	Deficiente
		2	Regular
		3	Buena
		4	Excelente
CALIFIQUE SU NIVEL DE HABILIDAD PARA ESCRIBIR LO SIGUIENTE:			
HAB_ARG	SU OPINIÓN SOBRE UN TEMA	1	No lo sé hacer
		2	Poco hábil
		3	Hábil
		4	Muy hábil
HAB_REP	EL REPORTE SOBRE UNA LECTURA	1	No lo sé hacer
		2	Poco hábil
		3	Hábil
		4	Muy hábil
HAB_HIS	UNA HISTORIA	1	No lo sé hacer
		2	Poco hábil
		3	Hábil
		4	Muy hábil
HAB_CAR	UNA CARTA/CORREO A UN FAMILIAR	1	No lo sé hacer
		2	Poco hábil

		3	Hábil
		4	Muy hábil
HAB_SOL	UNA SOLICITUD A UNA AUTORIDAD	1	No lo sé hacer
		2	Poco hábil
		3	Hábil
		4	Muy hábil
Sección 6. Exámenes de logro			
FAM_EXA	¿QUÉ TAN FAMILIARIZADO ESTÁ CON LOS EXÁMENES DE OPCIÓN MÚLTIPLE?	1	Nada familiarizado
		2	Poco familiarizado
		3	Familiarizado
		4	Muy familiarizado
INDICA LA FRECUENCIA CON LA QUE OCURREN LAS SIGUIENTES SITUACIONES CUANDO PRESENTA UN EXAMEN IMPORTANTE (EXAMEN FINAL, EXTRAORDINARIO, DE INGRESO, ETC.):			
FRE_SME	SALGO MAL EN LOS EXÁMENES PORQUE ME PONGO NERVIOSO	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_IEP	ENTRE MAS IMPORTANTE ES EL EXAMEN PEOR LO CONTESTO	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_ETP	POR ESTAR NERVIOSO EN UN EXAMEN ME TOMA MÁS TIEMPO ENTENDER LO QUE ME ESTÁN PREGUNTANDO	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_PPN	LA PRESIÓN DE TIEMPO CAUSA QUE ME PONGA NERVIOSO	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
FRE_BPE	ME BLOQUEO SI ALGUNA PREGUNTA EN UN EXAMEN ME CUESTA MUCHO TRABAJO	1	Nunca
		2	Algunas veces
		3	Frecuentemente
		4	Siempre
PRE_EXA1	¿PRESENTÓ EL EXAMEN NACIONAL DE INGRESO A LA EDUCACIÓN MEDIA SUPERIOR (EXANI - I) PARA INGRESAR AL BACHILLERATO?	1	Sí
		2	No
Sección 7. Entorno Social			
¿VIVE CON LAS SIGUIENTES PERSONAS?			
VIV_MAD	MADRE	1	Sí
		2	No
VIV_PAD	PADRE	1	Sí
		2	No
VIV_OTRO	OTROS FAMILIARES	1	Sí
		2	No
DUDA_PER	CUANDO TIENE UNA DUDA ACADÉMICA, ¿CUENTA CON UNA PERSONA QUE LE APOYE?	1	Sí
		2	No
ESCO_MAD	¿CUÁL ES EL MÁXIMO NIVEL DE ESTUDIOS ALCANZADO POR SU MADRE (AUNQUE HAYA	1	No estudió
		2	Primaria
		3	Secundaria

	FALLECIDO)?	4	Bachillerato
		5	Carrera técnica
		6	Licenciatura
		7	Posgrado (especialidad, maestría, doctorado)
		8	No lo sé
ESCO_PAD	¿CUÁL ES EL MÁXIMO NIVEL DE ESTUDIOS ALCANZADO POR SU PADRE (AUNQUE HAYA FALLECIDO)?	1	No estudió
		2	Primaria
		3	Secundaria
		4	Bachillerato
		5	Carrera técnica
		6	Licenciatura
		7	Posgrado (especialidad, maestría, doctorado)
		8	No lo sé
CUAN_LIB	¿CUÁNTOS LIBROS HAY EN SU CASA? (No considere revistas, periódicos o libros de texto).	1	Ninguno
		2	1 a 10
		3	11 a 25
		4	26 a 50
		5	51 a 100
		6	101 a 200
		7	201 a 500
		8	Más de 500
EN SU CASA, ¿CUENTA CON LOS SIGUIENTES BIENES Y SERVICIOS?			
SER_TELE	LÍNEA TELEFÓNICA	1	Sí
		2	No
SER_LAV	LAVADORA DE ROPA	1	Sí
		2	No
SER_REF	REFRIGERADOR	1	Sí
		2	No
SER_HOR	HORNO DE MICROONDAS	1	Sí
		2	No
SER_INTE	INTERNET	1	Sí
		2	No
SER_CABL	TELEVISIÓN DE PAGA (CABLEVISIÓN, SKY, ETC.)	1	Sí
		2	No
SER_TABL	TABLETA (IPAD, SAMSUNG GALAXY, ETC.)	1	Sí
		2	No
DEL SIGUIENTE LISTADO DE BIENES, SEÑALE CUÁNTOS HAY EN SU CASA.			
SER_DVD	REPRODUCTOR DE DVD	1	Ninguno
		2	1
		3	2
		4	3
		5	4 o más
BIEN_PC	COMPUTADORA	1	Ninguno
		2	1
		3	2
		4	3

		5	4 o más
SER_TV	TELEVISOR	1	Ninguno
		2	1
		3	2
		4	3
		5	4 o más
SER_AUTO	AUTOMÓVIL	1	Ninguno
		2	1
		3	2
		4	3
		5	4 o más
SER_BANO	BAÑOS COMPLETOS	1	Ninguno
		2	1
		3	2
		4	3
		5	4 o más
VAC_RM	EN LOS ÚLTIMOS 2 AÑOS, ¿CUÁNTAS VECES HA SALIDO DE VACACIONES DENTRO DE LA REPÚBLICA MEXICANA?	1	Ninguna
		2	1
		3	2
		4	3
		5	4 o más
EDO_REP	¿CUÁNTOS ESTADOS DE LA REPÚBLICA MEXICANA HA VISITADO COMO TURISTA?	1	Ninguno
		2	1
		3	2
		4	3
		5	4
		6	5
		7	6 o más
VI. CALIFICACIÓN DEL EXAMEN DE SELECCIÓN			
Calificación Global del examen de Selección			
POS_SEL	POSICIÓN ALCANZADA POR EL SUSTENTANTE EN EL EXAMEN DE SELECCIÓN	1-999999	
ICNE	CALIFICACIÓN EN ÍNDICE CENEVAL DEL EXAMEN DE SELECCIÓN	700-1300	
PERCEN	PERCENTIL DEL EXAMEN DE SELECCIÓN	0.00-100.00	
PORCECNE	% > CNE DEL EXAMEN DE SELECCIÓN	0.00-100.00	
PCNE	CALIFICACIÓN EN PORCENTAJE DE ACIERTOS DEL EXAMEN DE SELECCIÓN	0.00-100.00	
Calificación en porcentaje de aciertos del examen de Selección			

PRLM	CALIFICACIÓN DE RAZONAMIENTO LÓGICO MATEMÁTICO EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
PMAT	CALIFICACIÓN DE MATEMÁTICAS (SELECCIÓN) EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
PRV	CALIFICACIÓN DE RAZONAMIENTO VERBAL EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
PESP	CALIFICACIÓN DE ESPAÑOL EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
PTIC	CALIFICACIÓN DE TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
Calificación en índice Ceneval del examen de Selección			
IRLM	CALIFICACIÓN DE RAZONAMIENTO LÓGICO MATEMÁTICO EN ÍNDICE CENEVAL	700-1300	
IMAT	CALIFICACIÓN DE MATEMÁTICAS (SELECCIÓN) EN ÍNDICE CENEVAL	700-1300	
IRV	CALIFICACIÓN DE RAZONAMIENTO VERBAL EN ÍNDICE CENEVAL	700-1300	
IESP	CALIFICACIÓN DE ESPAÑOL EN ÍNDICE CENEVAL	700-1300	
ITIC	CALIFICACIÓN DE TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN EN ÍNDICE CENEVAL	700-1300	
VIII. CALIFICACIÓN DEL MÓDULO ESPECIAL DE INGLÉS			
Calificaciones del Módulo Especial de Inglés			
PMEI	CALIFICACIÓN DEL MÓDULO ESPECIAL DE INGLÉS EN PORCENTAJE DE ACIERTOS	0.00 - 100.00	
IMEI	CALIFICACIÓN DEL MÓDULO ESPECIAL DE INGLÉS EN ÍNDICE CENEVAL	700-1300	
DMEI	DICTAMEN DEL MÓDULO ESPECIAL DE INGLÉS	0	Sin Dictamen
		1	Elemental
		2	Satisfactorio
		3	Sobresaliente

La siguiente tabla muestra la definición de los atributos de las trayectorias:

Tabla 2. Descripción de los datos de las trayectorias

VARIABLE	DESCRIPCIÓN	VALORES Y ETIQUETAS
1.0	Tipo de plan de estudios:	1=Seis cuatrimestres,2=Despresurizado,
2.0	Promedio del bachillerato:	
3.0	Fecha de nacimiento:	
4.0	Sexo:	1=Mujer,2=Hombre,
5.0	Hablas alguna lengua indígena	1=Sí,2=No,
6.0	¿Provienes de alguna comunidad indígena?	1=Sí,2=No,
7.0	Estado civil:	1=Soltero,2=Casado o unión libre,
8.0	¿Tienes hijos?	1=Sí,2=No,
9.0	Indica la escolaridad máxima alcanzada por tus padres o tutores. Marca una sola opción en cada caso	1=Ninguna,2=Primaria completa,3=Secundaria completa,4=Bachillerato o equivalente completo,5=Estudios parciales de Licenciatura,6=Técnico Superior Universitario o Profesional Asociado completo,7=Licenciatura Completa,8=Posgrado completo,9=Lo ignoro,
9.1	Padre	
9.2	Madre	
10.0	¿Cuál es la ocupación actual de tus padres o tutores? Marca una sola opción en cada caso	1=Alto directivo (director, gerente, etc.),2=Empleado mando medio, superior (supervisor, jefe de departamento, etc.),3=Técnico de nivel operativo (mantenimiento, operador de equipo especializado, etc.),4=Empleado operativo en área administrativa (oficinista, atención al público, ventas, intendencia, etc.),5=Empleado operativo de servicios del estado (limpia, seguridad y vigilancia, servicios de agua, alcantarillado, alumbrado, etc.),6=Profesionista independiente (cuenta con consultorio, despacho, bufete, etc.),7=Profesor, maestro, docente (de cualquier nivel educativo),8=Propietario o copropietario de empresa a pequeña o mediana escala (de 21 a 100 empleados),9=Propietario o copropietario de microempresa: negocio o taller (de 1 a 20 empleados),10=Obrero (mano de obra en general: industrial, de la construcción, etc.),11=Campesino,12=Oficios y otras actividades (chofer, servicio doméstico, etc.),13=Migró al extranjero a trabajar (en cualquier actividad, descrita o no en la presente lista),14=Jubilado o pensionado,15=Labores del hogar exclusivamente,16=No trabaja,17=Lo ignoro,18=Finado,
10.1	Padre	
10.2	Madre	
11.0	¿En qué tipo de escuela realizaste tus estudios de bachillerato?	1=CONALEP,2=CBTIS,3=CETIS,4=CECYTE,5=Colegio de Bachilleres,6=Bachillerato a Distancia,7=Preparatoria Abierta,8=Preparatoria Pública,9=Preparatoria Privada,10=Otro,

11.1	Escribe el nombre de la escuela en la que realizaste tus estudios de bachillerato.	
12.0	¿Cuántos años empleaste para cursar tu bachillerato? Marca una sola opción	1=Dos años,2=Tres años,3=Más de tres años,
13.0	¿Trabajas?	1=Sí,2=No,
14.0	¿Económicamente, depende alguien de ti?	1=Sí,2=No,
15.0	¿Cuál es la principal razón por la que trabajas? Marca una sola opción	1=Pagarme mis estudios,2=Ayudar al gasto familiar,3=Sostener a mi familia,4=Tener independencia económica de mi familia,5=Adquirir experiencia laboral,
16.0	¿Cuántas horas trabajas a la semana? Marca una sola opción	1=10 hrs. o menos,2=De 11 a 20 hrs.,3=De 21 a 30 hrs.,4=De 31 a 40 hrs.,5=Más de 40 hrs.,
17.0	¿Tu trabajo tiene relación con la carrera que estás iniciando?	1=Sí,2=No,
18.0	Indica los medios de apoyo con los que cuentas para estudiar en casa. Marca una opción en cada uno de los renglones	1=SÍ,2=NO,
18.1	Espacio privado para estudiar	
18.2	Equipo de cómputo	
18.3	Internet	
18.4	Libros especializados	
18.5	Enciclopedias, diccionarios	
19.0	¿Qué importancia tienen tus estudios en las prioridades de tu familia? Marca una sola opción	5=Muy alta,4=Alta,3=Media,2=Baja,1=Muy baja,
20.0	Los recursos económicos con que cuentas para desarrollar tus actividades académicas son: Marca una sola opción	3=Óptimos,2=Suficientes,1=Insuficientes,
21.0	¿Cuál es el principal medio de transporte que utilizas para trasladarte a la Universidad? Marca una sola opción	1=Transporte propio de la Universidad,2=Transporte colectivo (Autobús, microbús, otro),3=Taxi,4=Vehículo particular,5=Caminando,
22.0	¿Cuánto tiempo haces regularmente para trasladarte de tu lugar de residencia a la Universidad? Marca una sola opción	1=Menos de ½ hr.,2=De ½ hr a 1 hr.,3=De 1 hr a 1 ½ hrs.,4=De 1 ½ hrs a 2 hrs.,5=Más de 2 hrs.,
23.0	Antes de hacer tu solicitud de	1=Sí,2=No,

	ingreso a la Universidad Tecnológica ¿Solicitaste admisión en otra institución de educación Superior?	
24.0	¿En qué tipo de institución de educación superior solicitaste admisión antes de tu ingreso a la Universidad?	1=Universidad Pública del Estado,2=Universidad privada,3=Instituto Tecnológico,4=Universidad Politécnica,5=Escuela Normal,6=Ninguna,7=Otra,
24.1	¿Cuál alternativa consideraste antes de ingresar a la Universidad Tecnológica?	
25.0	De no existir la Universidad Tecnológica ¿Qué decisión crees que habrías tomado al concluir tu bachillerato? Marca una sola opción	1=Inscribirme en otra institución de educación superior,2=Empezar a trabajar o continuar trabajando,3=Atender el negocio familiar,4=Dedicarme a las labores del hogar,5=Otras opciones de educación (inglés, cómputo, etc.),
26.0	En el proceso para que decidieras ingresar a la Universidad Tecnológica, ¿Qué tan importante fue la información que recibiste de los siguientes medios? Marca una opción en cada uno de los renglones	4=Muy Importante,3=Importante,2=Poco Importante,1=Nada Importante,
26.01	Televisión	
26.02	Radio	
26.03	Periódico	
26.04	Internet	
26.05	Volante, tríptico	
26.06	Información proporcionada por personal de la UT (vía telefónica, correo electrónico o ventanilla)	
26.07	Visita a la Universidad, expos y ferias	
26.08	Orientación vocacional en el bachillerato	
26.09	Recomendaciones de familiares, amigos o egresados de la UT	
26.1	Conversaciones con maestros del bachillerato	
27.0	En el proceso para que decidieras cursar tu carrera de TSU, ¿Qué importancia tuvieron los siguientes factores? Marca una opción en cada uno de los renglones	4=Muy Importante,3=Importante,2=Poco Importante,1=Nada Importante,0=No tuve conocimiento,
27.01	Que el modelo educativo está basado en competencias profesionales	
27.02	Duración de la carrera de TSU	
27.03	Realizar prácticas y estadías	

	en empresas	
27.04	Las oportunidades de empleo relacionadas con tu profesión	
27.05	El equipamiento de alta tecnología de la Universidad	
27.06	La preparación de los profesores	
27.07	Los programas de intercambio y movilidad estudiantil	
27.08	Los costos de inscripción y/o colegiaturas	
27.09	Los apoyos que ofrece la Universidad	
27.1	Que la Universidad ofrece Licencias Profesionales	
27.11	Que la Universidad ofrece Ingenierías	
28.0	De acuerdo con la carrera de TSU que estás iniciando, ¿En qué área laboral te gustaría desarrollarte? Marca una sola opción	1=En el sector público,2=Contratarme en una empresa privada,3=En el negocio de mi familia,4=Negocio propio / Ejercicio libre de la profesión,
29.0	Una vez que hayas concluido tu carrera de TSU ¿Cómo percibirás tu desarrollo profesional en comparación con tus padres? Marca una opción en cada uno de los renglones	5=Considerablemente mejor,4=Mejor,3=Similar,2=Inferior,1=Considerablemente inferior,
29.1	En prestigio Social	
29.2	En lo económico	
30.0	¿Cuál es tu principal objetivo al concluir tu carrera de TSU? Marca una sola una opción	1=Trabajar,2=Continuar una Ingeniería en la UT,3=Cursar una licenciatura o ingeniería en otra Institución de Educación Superior,
31.0	En comparación con otras opciones de educación superior ¿Cómo piensas que serán tus oportunidades al concluir tu carrera de TSU? Marca una opción en cada uno de los renglones	5=Considerablemente mejor,4=Mejor,3=Similar,2=Inferior,1=Considerablemente inferior,
31.1	Obtener empleo	
31.2	Trabajar en aquello para lo que te preparaste	
31.3	Mejor ingreso	
31.4	Oportunidades de desarrollo personal y profesional	
32.0	Durante el bachillerato ¿Realizabas alguna de las siguientes actividades para facilitar tu aprendizaje? Marca una opción en cada uno de los renglones	4=Siempre o casi siempre,3=Con bastante frecuencia,2=Con poca frecuencia,1=Casi nunca o nunca,
32.01	Buscabas información adicional a la que te	

	proporcionaban tus profesores	
32.02	Relacionabas los nuevos conocimientos con tus experiencias previas	
32.03	Organizabas distintos materiales para integrar tus trabajos	
32.04	Solicitabas que los maestros te aclararan tus dudas antes de realizar algún trabajo	
32.05	Realizabas ejercicios para reafirmar tu conocimiento	
32.06	Te planteabas preguntas a partir de la información que analizabas	
32.07	Discutías tus puntos de vista con maestros y compañeros	
32.08	Elaborabas esquemas, cuadros sinópticos, mapas conceptuales u otras estrategias para sintetizar la información	
32.09	Analizabas la eficacia de tus estrategias de estudio	
32.1	Buscabas nuevas formas de estudiar cuando alguna no te daba el resultado que esperabas	
32.11	Tomar apuntes en clase	
32.12	Colaborabas con tus compañeros de equipo para cumplir tareas	
32.13	Dirigías a tu equipo para el logro de las actividades académicas	
32.14	Tomabas decisiones en busca de mejores resultados en tus actividades académicas	
32.15	Negociabas acuerdos con tus compañeros de grupo	
32.16	Utilizabas equipo de cómputo para la elaboración de tus tareas	
32.17	Consultabas textos en inglés para apoyarte en la realización de tus tareas	
33.0	¿Qué tipo de fuentes de información empleabas en el bachillerato? Marca una opción en cada uno de los renglones	4=Siempre o casi siempre,3=Con bastante frecuencia,2=Con poca frecuencia,1=Casi nunca o nunca,
33.1	Bibliografía proporcionada por el profesor	
33.2	Bibliografía complementaria obtenida en biblioteca	

33.3	Buscadores de internet (google, yahoo, MSN, etc.)	
33.4	Sitios y enciclopedias online libres (wikipedia, monografias.com, etc.)	
33.5	Sitios web de instituciones educativas y revistas especializadas	
34.0	En promedio, ¿Cuántas horas a la semana dedicabas en el bachillerato a la preparación de tus clases y/o trabajos escolares? Marca una sola opción	1=De 1 a 5 horas,2=De 6 a 10 horas,3=Más de 10 horas,
35.0	¿Durante el último año del bachillerato, con qué frecuencia realizabas las siguientes actividades? Marca una opción en cada uno de los renglones	1=Nada,2=1-3 horas semanales,3=4-6 horas semanales,4=7-10 horas semanales,5=Más de 10 horas,
35.01	Convivir con la familia	
35.02	Convivir con amigos	
35.03	Practicar deporte	
35.04	Trabajar	
35.05	Ver televisión	
35.06	Lectura de actualidad (periódicos y revistas)	
35.07	Lectura de literatura (novela, cuento, poesía, etc.)	
35.08	Lectura de ocio (historietas, comics, revistas deportivas, de espectáculos, etc.)	
35.09	Practicar juegos de video	
35.1	Navegar y/o conversar por internet	
35.11	Escuchar música	
35.12	Practicar actividades artísticas	
36.1	En el aspecto cultural: ¿Participabas formalmente en alguna de las siguientes actividades cuando estabas en el bachillerato? Marca una opción en cada uno de los renglones	1=SÍ,2=NO,
36.11	Música	
36.12	Danza (Baile moderno, folklórico, etc.)	
36.13	Teatro	
36.14	Artes plásticas (Pintura, dibujo, escultura, fotografía, etc.)	
36.15	Cine club	
36.16	Círculo de lectura	
36.17	Otra actividad cultural	
36.171	Escribe la actividad cultural:	

36.2	En el aspecto deportivo: ¿Participabas formalmente en alguna de las siguientes actividades cuando estabas en el bachillerato? Marca una opción en cada uno de los renglones	1=SÍ,2=NO,
36.21	Futbol soccer	
36.22	Futbol rápido y 7	
36.23	Basquetbol	
36.24	Voleibol	
36.25	Atletismo (Carrera, Salto, Lanzamiento, etc.)	
36.26	Ajedrez	
36.27	Acondicionamiento físico (Pesas, Zumba, Aerobics, Jazz, Pilates, etc.)	
36.28	Otro deporte ¿Cuál?	
36.281	¿En qué otro deporte participabas?	
37.0	¿Participabas regularmente en alguna de las siguientes actividades cuando estabas en el bachillerato? Marca una opción en cada uno de los renglones	1=SÍ,2=NO,
37.1	Religiosa	
37.2	Política	
38.0	¿Durante el último año del bachillerato, con qué frecuencia asistías a los siguientes eventos? Marca una opción en cada uno de los renglones	1=Nada,2=1-4 veces en el año,3=5-8 veces en el año,4=9-12 veces en el año,5=Más de 12 veces en el año,
38.01	Cine	
38.02	Teatro	
38.03	Danza	
38.04	Exposiciones de pintura	
38.05	Presentación de libros	
38.06	Museos	
38.07	Conciertos de música clásica	
38.08	Conferencias	
38.09	Conciertos de música popular (rock, salsa, grupera, etc.)	
38.1	Ferias patronales o regionales	
38.11	Fiestas o reuniones con amigos	
38.12	Eventos deportivos	
39.0	¿Alguna vez has radicado fuera...?	2=No,1=SÍ,
39.1	De tu comunidad	
39.2	Del Estado	
39.3	Del país	
39.11	El motivo principal por el que	1=Familiar,2=Laboral,3=Escolar,

	te mudaste de tu comunidad fue	
39.21	El motivo principal por el que te mudaste de tu estado fue	1=Familiar,2=Laboral,3=Escolar,
39.31	El motivo principal por el que te mudaste de tu país fue	1=Familiar,2=Laboral,3=Escolar,

Usando librerías de WEKA en Java se evalúa el valor de un atributo mediante la medición de la relación de ganancia con respecto a la clase "promedio" para los atributos del CENVAL y las trayectorias. La siguiente figura muestra los atributos del CENEVAL con las trayectorias ordenados por importancia (ver Figura 5. Atributos del CENEVAL y trayectorias ordenados por importancia).

```
Atributos ordenados por importancia
0.4951 171 PROM_BAC
0.1651 11 EDO_PROC
0.1172 13 CIU_PROC
0.0992 18 BEC_DAC
0.0734 14 CVE_PROC
0.0719 12 NOM_PROC
0.0623 17 MOD_BAC
0.0485 53 11
0.0429 54 12
0.0384 15 ANO_BAC
0.0367 9 LI_MAD
0.0355 46 6
0.0342 48 8
0.0337 7 ANO_NAC
0.0274 45 5
```

Figura 5. Atributos del CENEVAL y trayectorias ordenados por importancia

A continuación se muestra la estadística descriptiva de los 4 atributos más importantes.

La siguiente figura muestra la estadística descriptiva para PROM_BAC (ver Figura 6. Estadística descriptiva PROM_BAC).

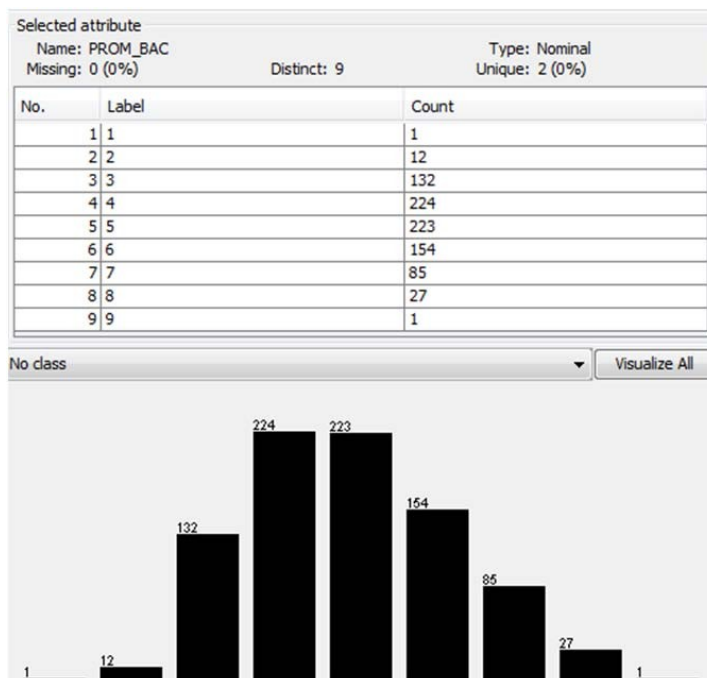


Figura 6. Estadística descriptiva PROM_BAC

La siguiente figura muestra la estadística descriptiva para EDO_PROC (ver Figura 7. Estadística descriptiva EDO_PROC).

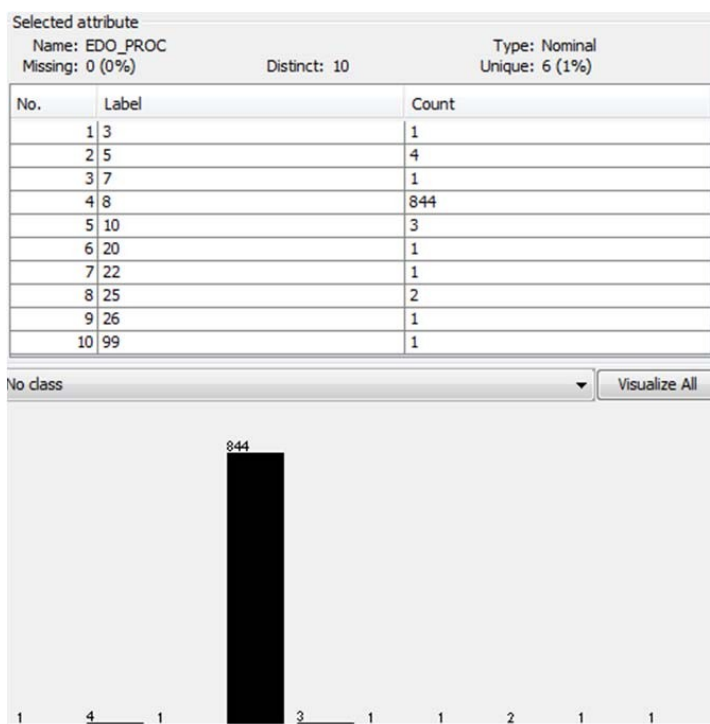


Figura 7. Estadística descriptiva EDO_PROC

La siguiente figura muestra la estadística descriptiva para CIU_PROC (ver Figura 8. Estadística descriptiva CIU_PROC).

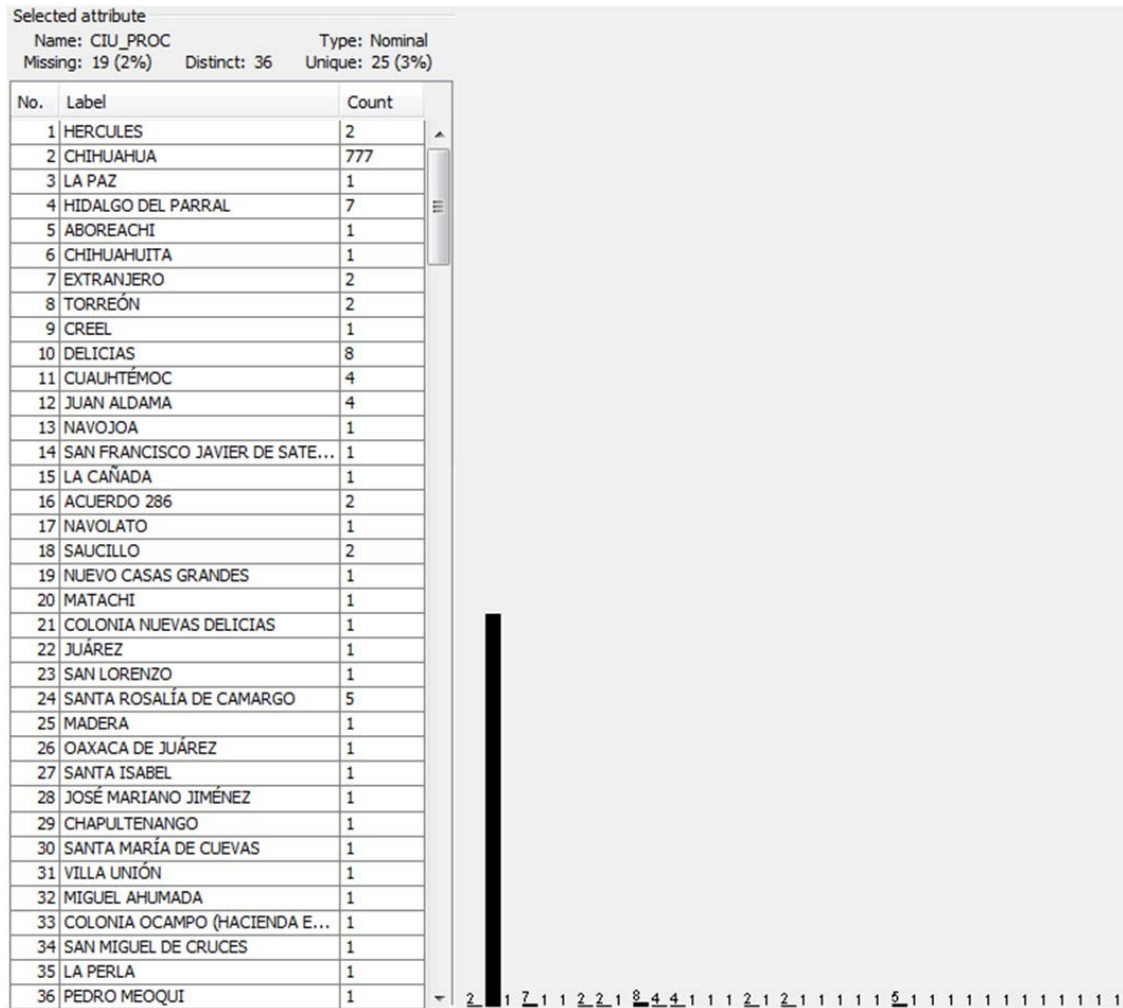


Figura 8. Estadística descriptiva CIU_PROC

La siguiente figura muestra la estadística descriptiva para BEC_DAC (ver Figura 9. Estadística descriptiva BEC_DAC).

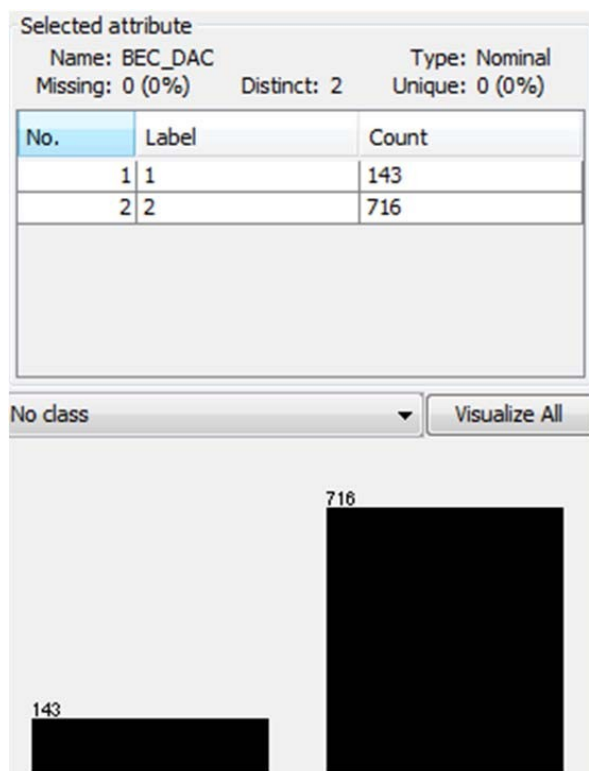


Figura 9. Estadística descriptiva BEC_DAC

4.2.4 Exploración de datos

Se tienen 4 fuentes de datos, las instancias y el nombre de las fuentes se muestran en la siguiente tabla:

Tabla 3. Fuentes de datos y sus instancias

Fuente	Instancias
CENEVAL	6424
Resultados de los cuatrimestres	25145
Propedéutico y Primer Cuatrimestre	2554
Trayectorias	1039

La minería de datos se procesa por medio de un modelo de clasificación, que ayudará a obtener el comportamiento predictivo para identificar a los alumnos con riesgo de desempeño escolar bajo.

4.2.5 Verificación de la calidad de los datos

Ya realizado el análisis y la exploración de los datos, se localizaron los siguientes conflictos:

- Instancias repetidas
- Campos vacíos o nulos.
- Datos mal capturados
- Errores de filas y columnas
- Tipo de dato incorrecto

Una vez encontrado todos estos problemas se determina que la gran mayoría de datos con conflictos se podían corregir, también que los demás datos estaban correctamente capturados y que se puede cubrir lo necesario para la obtención de los resultados requeridos en los objetivos del proyecto.

El pre-procesamiento realizado en los datos se detalla en la siguiente fase.

4.3 Preparación de los Datos

Esta etapa tiene como objetivo el transformar los datos obtenidos, que sean limpios, sin valores nulos y que estén listos para aplicarle las diferentes técnicas de minería de datos, además de unificar toda la información en vistas minables para encontrar los patrones que aciertan en la deserción estudiantil más fácilmente.

Para la seleccionar los datos adecuados, y realizar su análisis en WEKA se necesitó de un proceso el cual permitió obtener la vista minable de los datos:

- **Unificación de la información:** Para realizar la unión de los datos se elaboró un programa en AutoIt el cual fue programado para leer archivos excel, para concatenar la información de las celdas en base al identificador o identificadores asignados. Al programa también se le agregó la característica de poder cargar una cadena común para crear datos uniformes, ya que algunos de los datos del CENEVAL cambiaban entre versiones.

La información se guarda en dos repositorios. El primer repositorio contiene todos los datos académicos de los alumnos del primer cuatrimestre y todos los datos del CENEVAL. Estos datos se unieron mediante el nombre completo del alumno y se revisa que no existan nombres duplicados; El segundo repositorio contiene la unión de las trayectorias de investigación realizadas a los alumnos y los datos del CENEVAL. En el caso de las trayectorias se contó con la fecha de nacimiento además del nombre para realizar la unión entre ambos repositorios.

- **Limpieza de información:** Una vez realizada la unión de los datos se procedió a realizar la limpieza y corrección de los conflictos encontrados en la exploración de los datos. Algunos de los campos de datos numéricos se encontraban vacíos, así que fue necesario llenar esos campos con un valor desconocido "?" ya que no se pudo identificar un valor adecuado; se eliminaron los campos de la CURP, APE_PAT, APE_MAT, NOMBRE y MATRICULA ya que son identificadores del alumno, esto con el objetivo de proteger la privacidad del mismo; también se corrigieron conflictos con los datos en las celdas como duplicados y erróneos.
- **Datos innecesarios:** Se determinó que existían datos innecesarios como fue el caso de TIPO_EXA, APLI, FOLIO, entre otros, ya que contenían el mismo dato para cada una de las instancias o simplemente no contaban con información para el proceso de predicción.
- **Conversión de campos:** Muchos campos nominales y numéricos se exportaban erróneamente al WEKA ya que así se encontraban capturados en los archivos de excel, el ajuste final para los atributos se muestra en las siguientes tablas:

La siguiente tabla muestra el atributo y el tipo de atributo para los datos académicos de los alumnos y los datos del CENEVAL utilizados para el proceso de predicción.

Tabla 4. Descripción del tipo de atributo

ATRIBUTO	TIPO DE ATRIBUTO		
Sexo	NOMINAL	PROM_BAC	NOMINAL
Carrera	NOMINAL	BEC_DAC	NOMINAL
Cuatrimestre	NOMINAL	BEC_NEC	NOMINAL
Tipo plan	NOMINAL	BEC_HDA	NOMINAL
Turno	NOMINAL	HRS_TRAB	NOMINAL
FECHA_APLI_DIA	NOMINAL	EST_ALCA	NOMINAL
FECHA_APLI_MES	NOMINAL	NO_UNI	NOMINAL
FECHA_APLI_ANO	NOMINAL	SI_UNI	NOMINAL
IDENTIFICA	NOMINAL	SI_POSG	NOMINAL
DIA_NAC	NOMINAL	ESCO_MAD	NOMINAL
MES_NAC	NOMINAL	ESCO_PAD	NOMINAL
ANO_NAC	NOMINAL	CUAN_LIB	NOMINAL
LI_MAD	NOMINAL	SER_LAV	NOMINAL
LI_PAD	NOMINAL	SER_REF	NOMINAL
EDO_PROC	NOMINAL	SER_HOR	NOMINAL
NOM_PROC	NOMINAL	SER_INTE	NOMINAL
CIU_PROC	NOMINAL	SER_CABL	NOMINAL
CVE_PROC	NOMINAL	BIEN_PC	NOMINAL
ANO_BAC	NOMINAL	SER_TV	NOMINAL
REG_PROC	NOMINAL	SER_AUTO	NOMINAL
MOD_BAC	NOMINAL	SER_BANO	NOMINAL
		VAC_RM	NOMINAL
		PROMEDIO	NOMINAL

La siguiente tabla muestra el atributo y tipo de atributo para los datos de las trayectorias realizadas a los alumnos utilizados para el proceso de minería de datos.

Tabla 5. Descripción del atributo y su tipo

ATRIBUTO	TIPO DE ATRIBUTO		
Modalidad	NOMINAL	1	NOMINAL
Turno	NOMINAL	5	NOMINAL
familia	NOMINAL	6	NOMINAL
área	NOMINAL	7	NOMINAL
NombrePE	NOMINAL	8	NOMINAL
		9.1	NOMINAL
		9.2	NOMINAL



10.1	NOMINAL
10.2	NOMINAL
11	NOMINAL
12	NOMINAL
13	NOMINAL
14	NOMINAL
15	NOMINAL
16	NOMINAL
17	NOMINAL
18.1	NOMINAL
18.2	NOMINAL
18.3	NOMINAL
18.4	NOMINAL
18.5	NOMINAL
19	NOMINAL
20	NOMINAL
21	NOMINAL
22	NOMINAL
23	NOMINAL
24	NOMINAL
25	NOMINAL
26.01	NOMINAL
26.02	NOMINAL
26.03	NOMINAL
26.04	NOMINAL
26.05	NOMINAL
26.06	NOMINAL
26.07	NOMINAL
26.08	NOMINAL
26.09	NOMINAL
26.1	NOMINAL
27.01	NOMINAL
27.02	NOMINAL
27.03	NOMINAL
27.04	NOMINAL
27.05	NOMINAL
27.06	NOMINAL
27.07	NOMINAL
27.08	NOMINAL
27.09	NOMINAL
27.1	NOMINAL

27.11	NOMINAL
28	NOMINAL
29.1	NOMINAL
29.2	NOMINAL
30	NOMINAL
31.1	NOMINAL
31.2	NOMINAL
31.3	NOMINAL
31.4	NOMINAL
32.01	NOMINAL
32.02	NOMINAL
32.03	NOMINAL
32.04	NOMINAL
32.05	NOMINAL
32.06	NOMINAL
32.07	NOMINAL
32.08	NOMINAL
32.09	NOMINAL
32.1	NOMINAL
32.11	NOMINAL
32.12	NOMINAL
32.13	NOMINAL
32.14	NOMINAL
32.15	NOMINAL
32.16	NOMINAL
32.17	NOMINAL
33.1	NOMINAL
33.2	NOMINAL
33.3	NOMINAL
33.4	NOMINAL
33.5	NOMINAL
34	NOMINAL
35.01	NOMINAL
35.02	NOMINAL
35.03	NOMINAL
35.04	NOMINAL
35.05	NOMINAL
35.06	NOMINAL
35.07	NOMINAL
35.08	NOMINAL
35.09	NOMINAL

35.1	NOMINAL
35.11	NOMINAL
35.12	NOMINAL
36.11	NOMINAL
36.12	NOMINAL
36.13	NOMINAL
36.14	NOMINAL
36.15	NOMINAL
36.16	NOMINAL
36.17	NOMINAL
36.21	NOMINAL
36.22	NOMINAL
36.23	NOMINAL
36.24	NOMINAL
36.25	NOMINAL
36.26	NOMINAL
36.27	NOMINAL
36.28	NOMINAL
37.1	NOMINAL
37.2	NOMINAL

38.01	NOMINAL
38.02	NOMINAL
38.03	NOMINAL
38.04	NOMINAL
38.05	NOMINAL
38.06	NOMINAL
38.07	NOMINAL
38.08	NOMINAL
38.09	NOMINAL
38.1	NOMINAL
38.11	NOMINAL
38.12	NOMINAL
39.1	NOMINAL
39.2	NOMINAL
39.3	NOMINAL
39.11	NOMINAL
39.21	NOMINAL
39.31	NOMINAL
PROM_BAC	NOMINAL
PROMEDIO	NOMINAL

La unión de datos académicos de los alumnos con los datos del CENEVAL son un total de 1788 instancias con 46 atributos y la unión de los datos del CENEVAL con las trayectorias realizadas a los alumnos es un total de 859 instancias con 174 atributos.

También se realizó la conversión de las calificaciones de los resultados finales en las evaluaciones de los alumnos del primer cuatrimestre de numérico a nominal en base a los rangos proporcionados por los expertos de la UTCH. Los rangos de conversión se muestran en la siguiente tabla:

Tabla 6. Descripción de los datos

Desempeño	Rango
Autónomo	≥ 9.6
Destacado	≥ 8.6 y ≤ 9.59
Satisfactorio	≥ 8 y ≤ 8.59

Deficiente	< 8
------------	-----

Una vez realizado el pre-procesamiento de la información y la unión de los repositorios se procedió a trabajar con ellos para la creación de modelos de predicción. Estos repositorios fueron convertidos a CSV para poderlos cargar en WEKA y posteriormente guardarlos en formato ARFF.

4.4 Modelado

Esta etapa tiene como objetivo la búsqueda del mejor algoritmo de clasificación y de modelos que puedan predecir el desempeño académico de los alumnos. Para esto se carga la vista minable en la herramienta WEKA y se utilizan los algoritmos de clasificación J48, NaiveBayes y LibSVM. La siguiente figura muestra como se observan los datos cargados en WEKA (ver Figura 10. WEKA atributos e instancias).

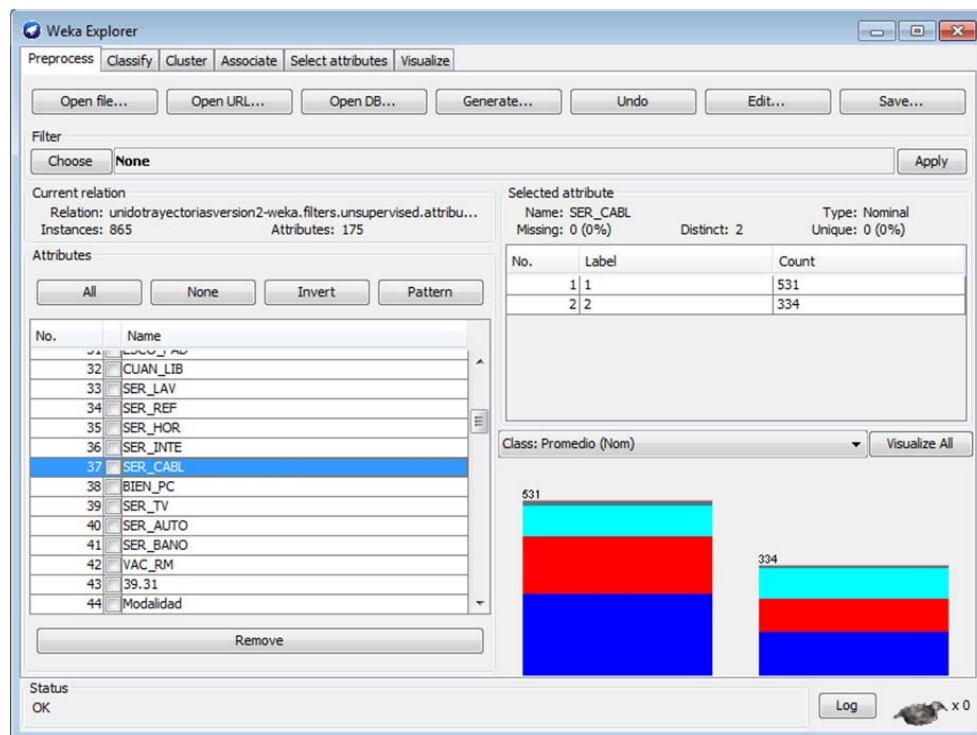


Figura 10. WEKA atributos e instancias

Para aplicar los algoritmos de clasificación en WEKA, se selecciona la pestaña Classify, se elige el botón choose de Classifier y dentro de la carpeta trees, bayes y rules se encuentran los algoritmos de clasificación J48, NaiveBayes y LibSVM respectivamente. por último se inicializa la minería de datos (ver Figura 11. Selección del algoritmo de clasificación).

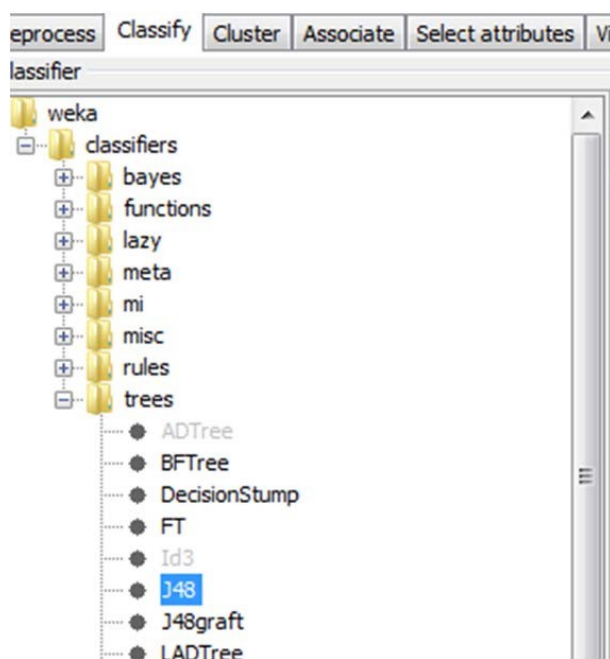


Figura 11. Selección del algoritmo de clasificación

4.4.1 Búsqueda del mejor algoritmo

La siguiente tabla muestra la precisión obtenida por cada uno de los algoritmos analizados utilizando las vistas minables de los datos del CENEVAL con los datos académicos de los alumnos y con los datos de las trayectorias realizadas a los alumnos.

Tabla 7. Descripción de los datos

Algoritmo	Precisión de los datos del CENEVAL con los datos académicos de los alumnos	Precisión de los datos del CENEVAL con los datos de las trayectorias realizadas a los alumnos
J48	85.46%	86.96%

NaiveBayes	81.94%	69.49%
LibSVM	79.08%	50.29%

Se eligió el algoritmo de clasificación J48 para la creación de modelos por ser el que mayor precisión tiene para predecir la clase PROMEDIO.

4.4.2 Algoritmo de clasificación J48

El modelo obtenido con el algoritmo de clasificación J48 para la vista minable de los datos del CENEVAL con los datos académicos de los alumnos obtuvo una precisión de 85.46% y un árbol de decisión de 8 reglas con un tamaño de 9 nodos (ver Figura 12. Árbol de decisión 1) el cual muestra principalmente que dependiendo el promedio de bachillerato (PROM_BAC) es el tipo de desempeño escolar que tendrá en su primer cuatrimestre.

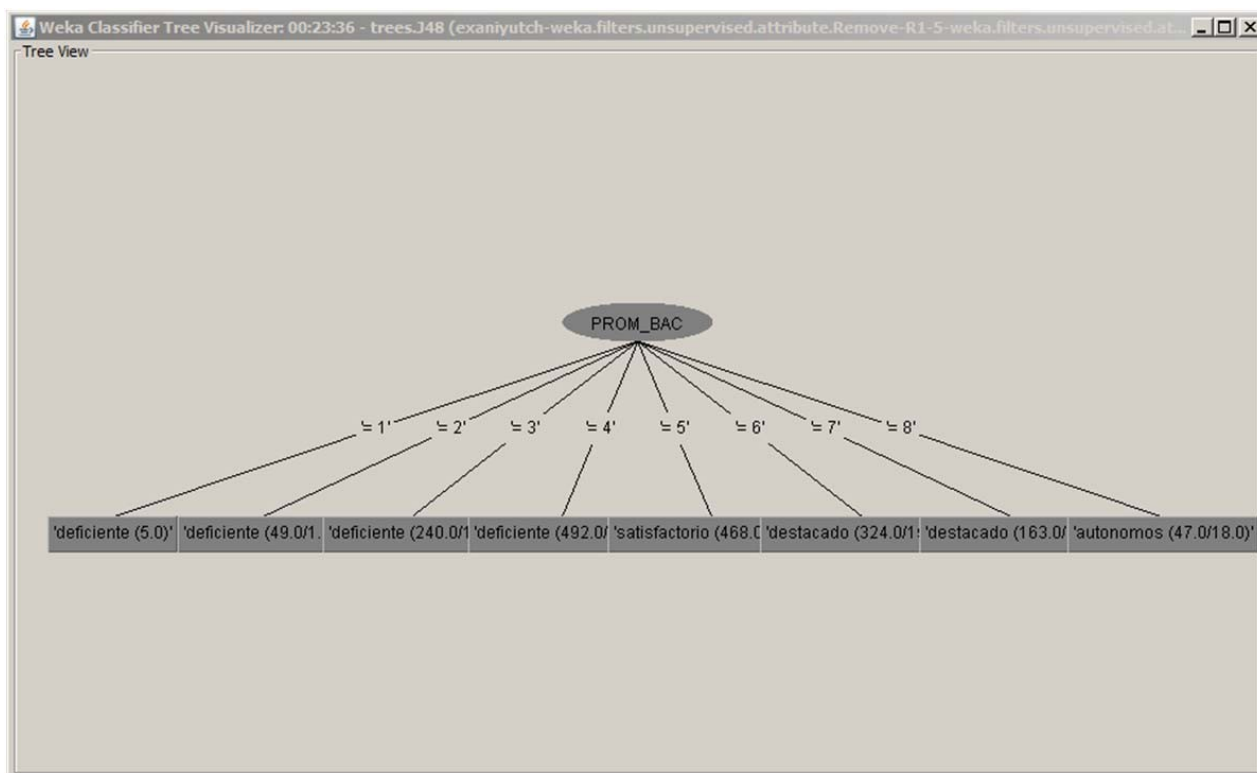


Figura 12. Árbol de decisión 1

El modelo obtenido para la vista minable de los datos del CENEVAL con los datos de las trayectorias realizadas a los alumnos obtuvo una precisión de 86.96% y un árbol de decisión de 19 reglas con un tamaño de 23 nodos (ver Figura 13. Árbol de decisión 2) el cual muestra otros factores importantes para el proceso de predicción los cuales son el promedio de bachilleres, un dato del aspecto cultural 36.16 (¿Participabas formalmente en “Cultura” cuando estabas en el bachillerato?), MOD_BAC (¿En qué modalidad obtuvo su certificado del bachillerato?) y 26.03 (Los recursos económicos con que se cuenta para desarrollar actividades académicas).

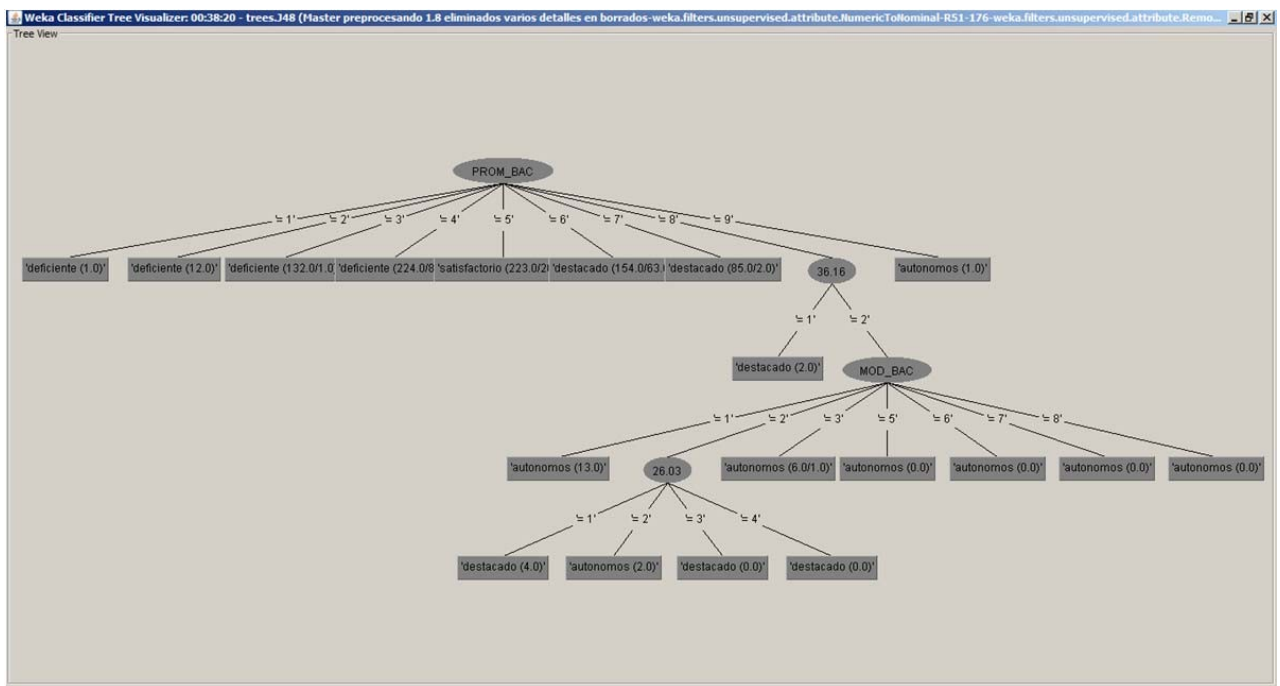


Figura 13. Árbol de decisión 2

4.4.2 Principio MDL

El principio MDL, es un método que nos permite seleccionar una hipótesis de acuerdo con su complejidad y la información que se requiere para describir la evidencia a partir de la hipótesis. Dicho de otro modo por medio del principio MDL se pueden obtener modelos más precisos y comprensibles. La fórmula del MDL es el factor f multiplicado por

la suma del número de condiciones del árbol más los errores obtenidos (Blanco, Lozoya, Roberto, & Blanco, 2015):

$$\text{MDL} = \text{factor} * \text{condiciones} + \text{errores}$$

La calidad de un árbol depende de dos aspectos, del error de predicción y del tamaño por ello se aplica el principio MDL partiendo de la comprensión más la precisión del modelo obtenido con el algoritmo J48.

Para utilizar el principio MDL se realiza un programa en Java usando Netbeans en el cual se carga la vista minable del CENEVAL con las trayectorias realizadas a los alumnos con el objetivo de generar varios modelos y obtener el modelo de clasificación con el mejor MDL.

El objetivo de crear varios modelos es para que los expertos de la UTCH seleccionen el modelo que más les sirviera, que sea útil y novedoso. A continuación se muestran los modelos obtenidos considerando el principio MDL.

El siguiente modelo se realiza con la vista minable de la unión del CENEVAL con las trayectorias realizadas a los alumnos; tiene un MDL igual a 192, una precisión de 88.13% y un árbol de decisión de 9 hojas con un tamaño de 10. El atributo que se considera en el MDL para el proceso de predicción es el promedio de bachilleres (PROM_BAC). La siguiente figura muestra el árbol de decisión generado en WEKA (ver Figura 14. Árbol de decisión 3).

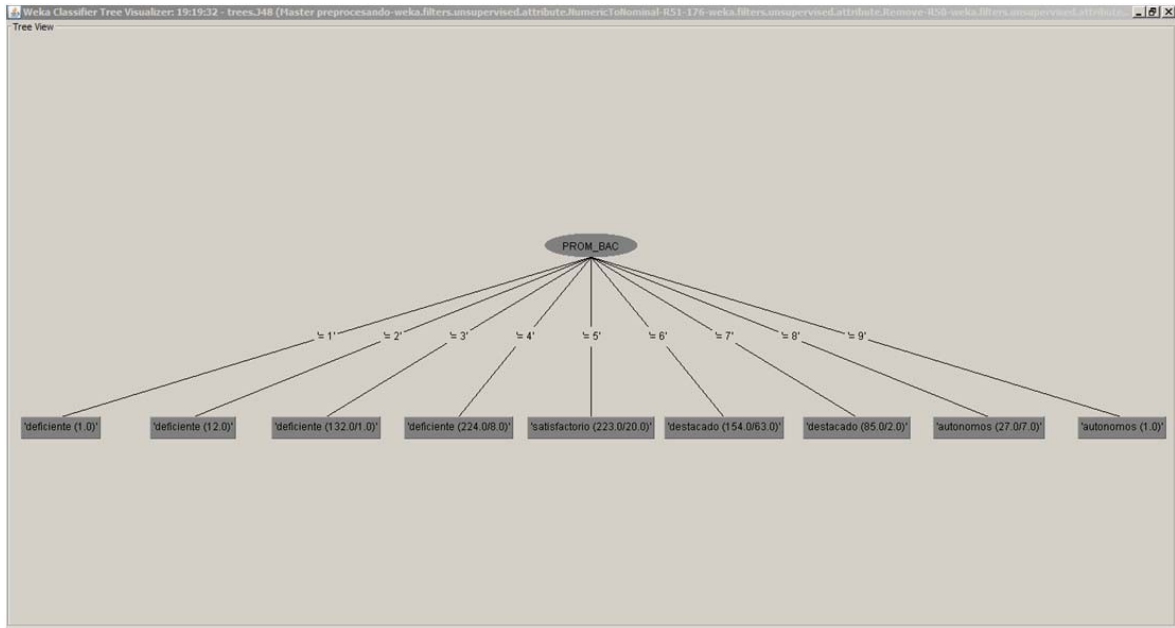


Figura 14. Árbol de decisión 3

El siguiente modelo se realizó con el repositorio de la unión del CENEVAL con las trayectorias realizadas a los alumnos; tiene un MDL igual a 215, una precisión de 88.94% y un árbol de decisión de 11 hojas con un tamaño de 13. Los atributos que se consideraron en el MDL para el proceso de predicción son el promedio de bachilleres (PROM_BAC), el estado de procedencia (EDO_PROC), nombre de la institución donde concluyó el nivel medio superior (NOM_PROC), nombre de la ciudad donde se ubica la institución en la que concluyó el nivel medio superior (CIU_PROC), clave de la institución en la que concluyó sus estudios de bachillerato (CVE_PROC), modalidad en la que se obtuvo su certificado de bachillerato (MOD_BAC), obtuvo beca por alto desempeño académico (BEC_DAC), tiempo empleado para cursar el bachillerato (12) y el tipo de escuela en donde se realizó los estudios de bachillerato (11). La siguiente figura muestra el árbol de decisión generado en WEKA (ver Figura 15. Árbol de decisión 4).

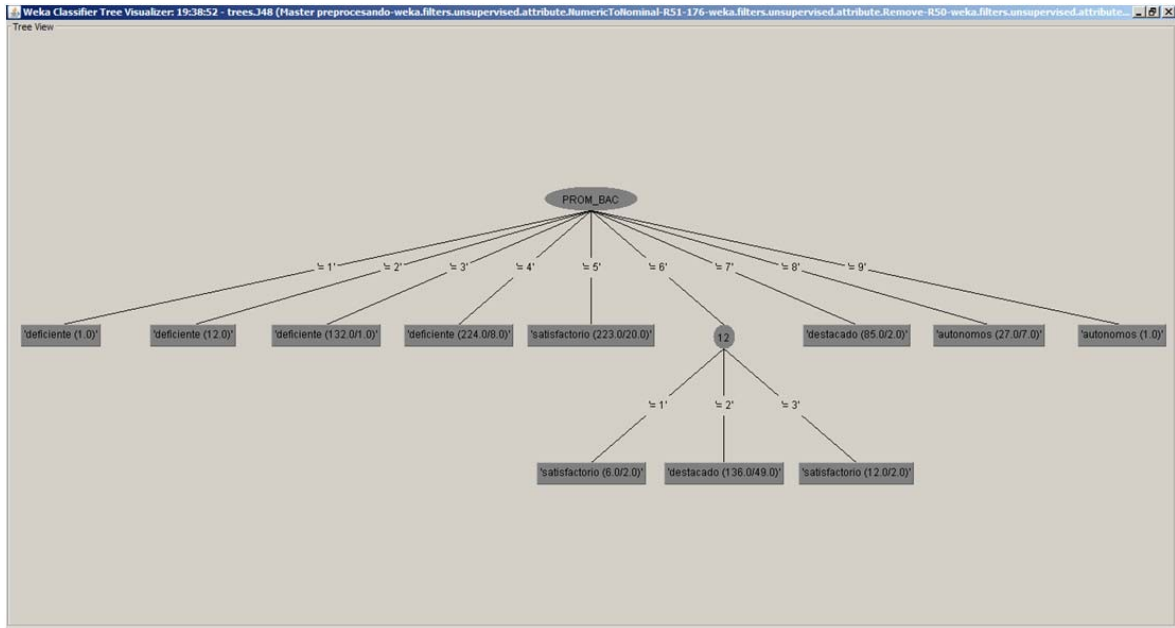


Figura 15. Árbol de decisión 4

El siguiente modelo se realizó con el repositorio de la unión del CENEVAL y los resultados académicos de los alumnos; tiene un MDL igual a 340, una precisión de 85.90% y un árbol de decisión de 14 hojas con un tamaño de 16. Los atributos que se consideraron en el MDL para el proceso de predicción fueron el promedio de bachilleres (PROM_BAC) y la modalidad en la que se obtuvo su certificado de bachillerato (MOD_BAC). La siguiente figura muestra el árbol de decisión generado en WEKA (ver Figura 16. Árbol de decisión 5).

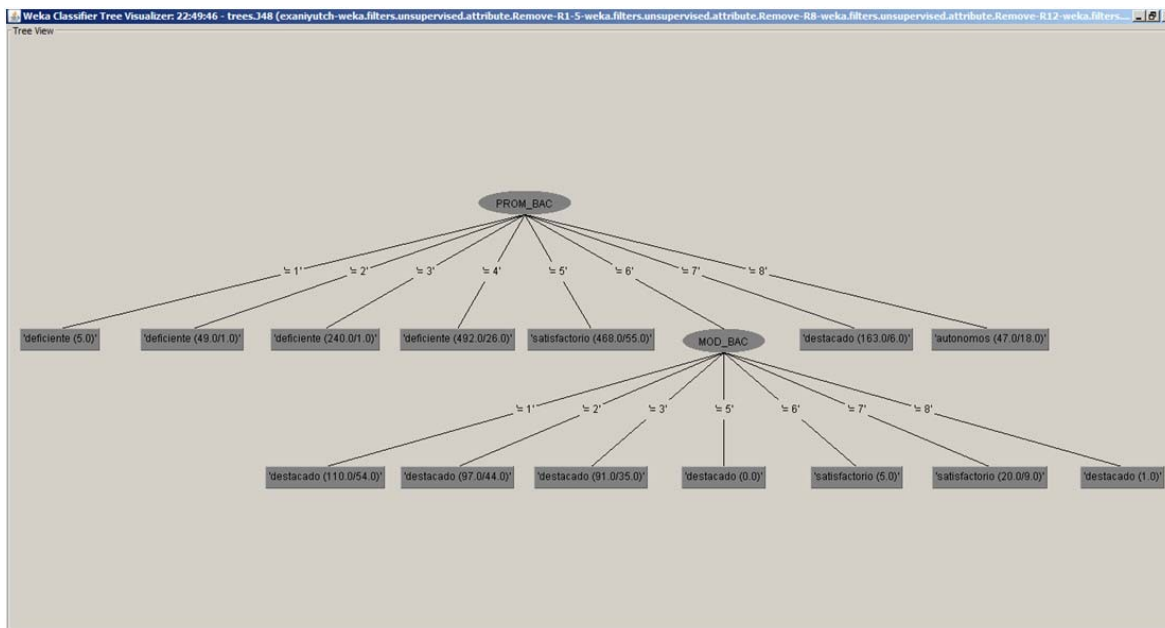


Figura 16. Árbol de decisión 5

El siguiente modelo se realizó con el repositorio de la unión del CENEVAL y los resultados académicos de los alumnos; tiene un MDL igual a 340, una precisión de 85.46% y un árbol de decisión de 8 hojas con un tamaño de 9. El atributo que se consideró en el MDL para el proceso de predicción fue el promedio de bachilleres (PROM_BAC). La siguiente figura muestra el árbol de decisión generado en WEKA (ver Figura 17. Árbol de decisión 6).

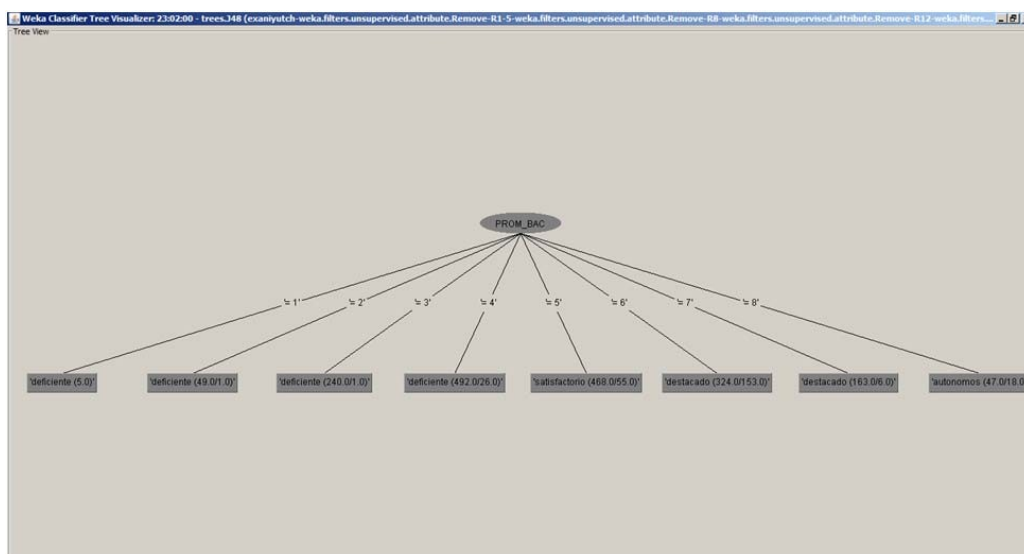


Figura 17. Árbol de decisión 6

4.5 Evaluación

Esta etapa tiene como objetivo evaluar los modelos obtenidos en la fase de modelado y la selección de un modelo por medio de los expertos de la UTCH. Este modelo se utiliza con el sistema de predicción.

Se realizaron 4 modelos con el principio MDL, 2 de ellos correspondientes a la unión del CENEVAL con las trayectorias y los otros 2 a la unión del CENEVAL con los resultados académicos de los alumnos. A continuación se muestra la tabla comparativa de los 4 modelos utilizando el principio MDL.

Tabla 8. Descripción de los datos

	Precisión	Hojas	Nodos
MDL1: CENEVAL y resultados académicos de los alumnos	85.90%	14	16
MDL2: CENEVAL y resultados académicos de los alumnos	85.46%	8	9
MDL3: CENEVAL y trayectorias	88.13%	9	10
MDL4: CENEVAL y trayectorias	88.94%	11	13

Los árboles de decisión de la figura 17 y de la figura 14 se parecen mucho, solo difieren en que el de la figura 14 cuenta con el rango de 9 en la calificación del bachillerato y el de la figura 17 no. Esto se debe a que al momento de realizar la unión con los datos proporcionados por la UTCH los alumnos del CENEVAL que obtuvieron una calificación de tipo 9 no se encontraban en los resultados académicos de los alumnos del primer cuatrimestre pero si en las trayectorias.

Otro detalle a considerar es que el modelo de la unión del CENEVAL con las trayectorias generan mejores resultados incrementando la precisión alrededor de un 3.5%. No obstante los expertos de la UTCH mencionan que tienen dificultad obteniendo las trayectorias que se les realizan a los alumnos por tal motivo los expertos seleccionaron el modelo de la unión del CENEVAL con los resultados académicos de los alumnos; el cual tiene como atributos importantes el promedio de bachilleres (PROM_BAC) y la modalidad en la que se obtuvo su certificado de bachillerato (MOD_BAC).

4.6 Implementación

Esta etapa tiene como objetivo el implementar una aplicación que tenga una interfaz intuitiva y fácil de usar por los docentes o usuarios finales encargados de la gestión estudiantil. Esta aplicación hará uso de la información recabada y procesada, con el objetivo de que se pueda detectar de manera más fácil si un alumno tiene probabilidad de tener un desempeño escolar bajo.

Obtenido el modelo se procede a su implementación, para ello se programó una aplicación para la predicción.

A continuación se muestra la aplicación una vez cargado el modelo y un ejemplo de una prueba de predicción (ver Figura 18. Sistema de predicción y Figura 19. Ejemplo de predicción).

Sistema de Predicción Para: Universidad Tecnológica de Chihuahua

Nombre del Alumno: Pedro Morales

PROM_BAC: 6

Predecir... Rangos

Figura 18. Sistema de predicción

Mensaje

Se predice que el alumno Pedro Morales tendra un Promedio destacado en su primer cuatrimestre con una calificación entre 8.6 y 9.59

Aceptar

Figura 19. Ejemplo de predicción

También se incluyó en el programa los rangos del promedio de bachilleres. La siguiente figura muestra los rangos del promedio de bachilleres (ver Figura 20. Rangos del promedio de bachilleres).



Figura 20. Rangos del promedio de bachilleres

CAPÍTULO V. CONCLUSIÓN

Utilizando WEKA se experimentó con los algoritmos de clasificación J48, LibSVM y NavieBayes, permitiendo concluir que el mejor algoritmo para nuestro modelo de predicción es el J48 ya que contiene la más alta precisión de los 3 con un 85.46% para los datos del CENEVAL con los datos académicos de los alumnos y un 86.96% para los datos del CENEVAL con las trayectorias realizadas a los alumnos.

La selección del mejor modelo de predicción es un proceso difícil. Regularmente se utiliza solamente la precisión. En esta investigación se incluyó además de la precisión, la comprensibilidad. Por lo que es necesario considerar las propiedades subjetivas del conocimiento. Por tal motivo se utilizó al experto para validar el conocimiento aprendido por medio de la utilidad y novedad para obtener el sistema de predicción de más alta calidad. El mejor modelo seleccionado por el experto fue el MDL2: CENEVAL y resultados académicos de los alumnos, con un 85.46% de precisión y un árbol de decisión de 8 hojas con un tamaño de 9 nodos que tiene como mejor atributo el promedio de bachilleres (PROM_BAC), ya que los datos necesarios para la predicción de este modelo siempre se pueden conseguir por el experto y los datos de las trayectorias no.

El sistema de predicción creado para la UTCH brinda una expectativa al alumnado de su posible desempeño escolar en su primer cuatrimestre, haciendo posible que estudiante y la institución puedan tomar medidas preventivas para evitar circunstancias de bajo rendimiento académico.

Queda comprobada la hipótesis de que si es posible estimar el desempeño académico de los alumnos en su primer cuatrimestre en la UTCH mediante un sistema de predicción.

REFERENCIAS

- Amaya, K., Barrientos, E., & Heredia, J. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos.
- AutoIt. (2016). <https://www.autoitscript.com/site/autoit/>.
- Blanco, R., Lozoya, L., Roberto, L. J., & Blanco, H. (2015). Maximizar La Calidad Objetiva De Los Árboles De Decisión. *CIPITECH*.
- Bonsón, E., Escobar, T., & Zamora, M. (s.f.). *Sistemas De Inducción De Árboles De Decisión: Utilidad En El Análisis De Crisis Bancarias*. Obtenido de <http://ciberconta.unizar.es/docencia/intelig>
- CENEVAL. (2015). *Centro Nacional De Evaluación Para La Educación Superior*. Obtenido de <http://www.ceneval.edu.mx/>
- Figuerroa, C. (2004). *Sistemas de Evaluación Académica*. El Salvador: Editorial Universitaria.
- Formia, S., Lanzarini, L., & Hasperué, W. (2013). Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 92-98.
- García, R. (2012). Minería De Datos Para La Toma De Decisiones E Inteligencia De Negocios : Aplicaciones En La Mercadotecnia.
- Garibay, A. P., & Galicia, S. (2015). Clasificación automática de la orientación semántica de opiniones mediante características lingüísticas. *UNAM*.
- Gómez, A. A. (2013). Modelo De Detección De Estudiantes Excluidos En Carreras De Ingeniería Utilizando Minería De Datos. *Universidad Francisco de Paula Santander Ocaña*.
- Greiner, L. (2014). Bases de Datos.
- Montes, C., & Lerner, J. (2010). Rendimiento Académico de los estudiantes de pregrado de la Universidad EAFIT. Perspectiva cuantitativa.
- Paute, D. J. (2016). Análisis Y Clasificación De Información Mediática Electoral Utilizando Minería De Texto. *Universidad del País Vasco*.
- Rokach, L., & Maimon, O. (2015). *Data Mining With Decision Trees Theory And Applications*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Romero Morales, C., Márquez Vera, C., & Ventura Soto, S. (2012). Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos. *IEEE-Rita*.
- Suca, C., Córdova, A., Condori, A., Cayra, J., & Sulla, J. (2016). Comparación De Algoritmos De Clasificación Para La Predicción De Casos De Obesidad Infantil. *Universidad Nacional de San Agustín*.
- Toledo, J. J., Romero, A. C., & Pereira, R. T. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Universidad de Nariño*.
- Valero, S., García, M., & Salvador, A. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos.



CURRICULUM VITAE

Nombre: Luis Alberto Lozoya Trevizo
Edad: 28 años
Estado civil: Soltero
Domicilio: Lucio Cabañas #1002
Ciudad: Chihuahua, Chih.
Teléfono: (614) 1307086 cel.
Correo electrónico: lozoyaluis@hotmail.com



OBJETIVO:

Esforzarme con los requerimientos que la empresa solicite, cumplir eficientemente cada una de las actividades asignadas. También adquirir nuevas habilidades y nuevos conocimientos que me permitan tener un mejor desempeño.

FORMACION ACADEMICA:

Técnico en Electrónica(Titulado)

CBTis 122

Ingeniero en Sistemas Computacionales(Titulado)

Instituto Tecnológico de Chihuahua II

IDIOMAS:

- Español (lenguamaterna).
- Inglés (estudiada en el extranjero).

EXPERIENCIA PROFESIONAL:

Ford Productos Automotrices SA de CV.- Labore durante un año y medio:

1. Programando principalmente en php, html y javascript, dando mantenimiento, actualizando módulos y templates del sistema que la empresa tiene, además del desarrollo de páginas web.
2. En la instalación y configuración de un servidor para interconectarlo con el servidor principal de la empresa.
3. Dando mantenimiento de equipo de computo, además de soporte a usuarios.

Dirección de Vialidad y Protección Civil.- Laboré alrededor de 1 año, con el proyecto de reestructuración de red y análisis del ancho de banda, donde destacaron los siguientes beneficios:

1. Evitar problemas al buscar nodos.
2. Mayor control de los usuarios en los equipos.
3. Mayor efectividad en la transmisión de datos.
4. Implementación de seguridad en el acceso a los dispositivos de interconexión.

HABILIDADES Y CUALIDADES:

Lenguajes de programación:

- | | |
|--------------------------|----------|
| • Delphi | • HTML |
| • java(Netbeans) | • PHP |
| • Microsoft VisualStudio | • Autoit |

Bases de datos:

- MySQL
- SQL Server