

Capítulo 4 Identificación de factores de riesgo que causan la deserción de alumnos que estudian a distancia por causa del COVID19 usando técnicas de minería de datos

Chapter 4 Identification of risk factors that cause the dropout of students studying at a distance due to COVID19 using data mining techniques

REYES-NAVA, Adriana†*, GIL-ANTONIO, Leopoldo y ANTONIO-VELÁZQUEZ, Juan Alberto

Tecnológico de Estudios Superiores de Jocotitlán, División de Ingeniería en Sistemas Computacionales.

ID 1^{er} Autor: *Adriana, Reyes-Nava* / **ORC ID:** 0000-0002-4440-909X

ID 1^{er} Coautor: *Leopoldo, Gil-Antonio* / **ORC ID:** 0000-0002-7445-9426

ID 2^{er} Coautor: *Juan Alberto, Antonio-Velázquez* / **ORC ID:** 0000-0003-3052-3171

DOI: 10.35429/H.2021.11.62.72

A. Reyes, L. Gil y J. Antonio

*adriana.reyes@tesjo.edu.mx

A. Ledesma (Coord.) Ciencias de la Ingeniería y Tecnología. Handbooks-TX-©ECORFAN-México, 2021.

Abstract

The students dropout in educational institutions of higher level in México its a problem given there are high rates of students in risk. A solution for this problem is classification of risk factors that cause this problem. These factors are mainly classified as: academic, socioeconomic and even family. In this work, the first phase of the study is considered, which consists of exploring the data through the use of the *A priori algorithm of association rules* through the use of data mining. For this, a Database (DB) will be used that contains information collected through the application of forms for students of the career. Once the information is obtained from the database, it will be pre-processed, eliminating redundant or missing data and generating new samples from distribution of collected data. Subsequently, data mining is applied, using Weka software, to interpret the patterns of each risk factor.

Data-Mining, School dropout, Association rules, Weka

1. Introducción

El COVID19 o también conocido como (SARS CoV-2), hoy en día es una de las enfermedades que está causando pérdidas humanas alrededor del mundo y sobre todo en México. Esta pandemia ha causado que los estudiantes en todos los niveles educativos tomen sus clases en línea, lo cual ha ocasionado que ellos sufran de diversos factores que les impide seguir estudiando en esta modalidad (Garcés Fuenmayor & Mora Bolaños, 2020), es por eso que las entidades educativas buscan estrategias para que la población estudiantil permanezca y no disminuya debido a diversos factores que causan esta disminución (Pérez Zuñiga, Martínez García, & Mena Hernández, 2021). Por otro lado, en (Rincón, Sir, & Suárez Castrillón, 2020), se menciona que, en la Universidad Francisco de Paula Santander Ocaña, recabaron la información de los estudiantes inscritos de los primeros y segundos semestres de los años 2019 y 2020. Los resultados revelaron que el índice de deserción en la época de pandemia fue del 8.5% a pesar de que los profesores y el gobierno ofrecieron ayuda moral y económica a los estudiantes para que no abandonaran sus estudios.

En (Venegas Morales & Rubio Hernández, 2021), realizan un estudio en una muestra de población estudiantil y docente en los meses de junio y julio del 2020 y concluyen que hay una similitud en las dificultades sociales, emocionales y familiares acarreaban problemas que hacían que los jóvenes desertaran. Sin embargo, en (Sandoval Elías, 2021), identifican cual es la problemática de las causas que originan la continuidad y no continuidad en la formación académica de estudiantes de nivel superior en la Facultad de humanidades de la Escuela de Estudios de Posgrado, esta problemática radica en que la nueva modalidad virtual ha influido para que muchos estudiantes se retiren de las clases por no contar con las herramientas necesarias y/o recursos económicos para mantener una conectividad y realizar actividades o tareas económicas.

Por otro lado, el uso de técnicas estadísticas ha ayudado a analizar esta problemática, como es el caso de la investigación realizada por (Benavides-Lucksic, Villena-Presentación, & Andía-Valencia, 2020), ellos determinaron como indicador la cantidad de alumnos matriculados en un semestre determinado y compararlo con otro semestre similar, para esto utilizaron un método estocástico (cadenas de Markov), con el cual calcularon el muestreo estadístico para obtener como resultado las diferentes matrices que permiten predecir eventos futuros de deserción. En (Avila Pérez, 2021), se hace un estudio sobre la deserción de alumnos en la UNAD, aplican técnicas de analítica de datos a la información que se almacena como producto de los procesos académicos, ellos aplican técnicas de minería de datos para pronosticar y crear un modelo de predicción que pueda evitar que los alumnos deserten ante cualquier circunstancia. Por su parte (Contreras Bravo, Fuentes López, & González Guerrero, 2020), realizan un análisis de los cambios que están sufriendo el sector educativo debido a la adopción de tecnologías y plataformas que han cambiado la forma de tomar clases, para esto se utilizan técnicas basadas en machine learning que permitan la toma de decisiones y así tomar decisiones oportunas para evitar la deserción debido a problemas psicológicos causados por la frustración y ansiedad debido al encierro de los estudiantes por la pandemia del COVID19. Es por eso que en este trabajo se muestran los factores que influyen en la deserción escolar de alumnos de la carrera de Ingeniería en Sistemas Computacionales del Tecnológico de Estudios Superiores de Jocotitlán, durante el periodo de contingencia, se muestran los hechos que han ocurrido a lo largo del tiempo en la educación superior en todo el país, así como la situación actual del Tecnológico por el desempeño de los alumnos en las clases en línea.

Para esto se considera la primera fase del estudio, que consiste en la exploración de los datos mediante el uso del algoritmo *A priori de reglas de asociación* mediante las etapas de minería de datos. Se considera para entrenar el algoritmo una base de datos (BD), que contiene información recabada mediante aplicación de formularios a los alumnos de la carrera de ingeniería en sistemas computacionales, estos formularios contienen preguntas específicas relacionadas a los problemas que los estudiantes enfrentan en la modalidad en línea. Una vez obtenida la información de la BD, ésta será preprocesada, eliminando datos redundantes o faltantes. Posteriormente se aplican las etapas de minería de datos, mediante el uso del software Weka, con lo cual se podrá interpretar los patrones de cada factor de riesgo.

El presente trabajo se organiza en 5 etapas, donde se tratan los puntos relacionados con el tema de investigación, organizados de la siguiente manera:

- Etapa 1. Introducción, se presenta el enfoque del trabajo, hace referencia al objetivo, la problemática a resolver, justificación e hipótesis del mismo, así como el impacto científico y tecnológico que tendrá el trabajo.
- Etapa 2. Panorama actual y fundamento teórico, aquí se presentan los trabajos relacionados a este tema de investigación, así como los conceptos fundamentales de minería de datos.
- Etapa 3. Marco metodológico, en este capítulo se muestran los pasos a seguir para llevar a cabo la experimentación del proceso de minería de datos.
- Etapa 4. Resultados, se muestran los resultados obtenidos con base en la experimentación realizada.
- Etapa 5. Conclusiones, se presenta la conclusión a la cual se llegó con base en los resultados obtenidos y se muestra el trabajo futuro que se planea realizar.

2. Panorama actual y fundamento teórico

En el estado del arte se han analizado las diferentes investigaciones que tienen como objetivo identificar cuáles son los factores de riesgo que ocasionan que un alumno deserte, haciendo uso de métodos de inteligencia artificial como minería de datos y aprendizaje automático. En la mayoría de los trabajos se hace un énfasis en la educación presencial, sin embargo, actualmente por la situación que se vive es necesario analizar los factores de riesgo generados por los alumnos que estudian en línea, estos factores son almacenados en forma de datos y al final generan resultados que pueden cambiar por diferentes situaciones de estrés.

La educación y la formación profesional en el nivel medio superior es menos frecuente en México que en otros países que pertenecen a la OCDE. En promedio, el 35% de los estudiantes de educación media superior están inscritos en programas de formación profesional en México, muy por debajo del promedio de 42% que corresponde a los países que pertenecen a la OCDE.

Los estudiantes que reciben formación profesional en la educación media superior obtienen el beneficio de tener altos niveles de acceso a la educación superior en México. En México, el 97% de los estudiantes que reciben formación profesional en la educación media superior están inscritos a programas que ofrecen acceso directo a la educación superior, lo cual es una cifra superior al promedio de 70% de la OCDE. El nivel de educación superior en México sigue siendo bajo. Solo el 24% de los jóvenes de entre 25 y 34 años de edad obtuvieron un título de estudios superiores en el 2019 en México, en comparación con el promedio de 45% en los países que pertenecen a la OCDE.

En el 2019, el 2% de los jóvenes que tenían entre 25 y 34 años de edad en México tenía un título de formación profesional de nivel medio superior como su nivel educativo más alto, mientras que el 26% tenía uno general. La tasa de empleo en los adultos más jóvenes que tienen una educación profesional media superior tiende a ser más alta que la tasa de empleo de los que tienen una certificación general al mismo nivel (con un promedio de 9 puntos porcentuales en los países que pertenecen a la OCDE). México sigue este patrón, ya que el 76% de los jóvenes que tienen entre 25 y 34 años de edad con un certificado de educación media superior tienen un empleo, en comparación con el 71% de los que tienen un certificado general

En México, la edad promedio de quienes ingresaron por primera vez a la educación superior en el 2018 fue de 21 años, lo cual es inferior al promedio de 22 años de la OCDE. Los factores estructurales, como los procedimientos de admisión, la edad usual en que los estudiantes se gradúan de la educación media superior o las percepciones culturales sobre el valor de las experiencias profesionales o personales que se obtienen fuera de la educación pueden explicar las diferencias en la edad promedio de ingreso a la educación superior entre los distintos países. (Panorama de la educación 2020, (2021))

En (Miranda, Mauricio A, & Guzmán, Jheser. (2017)) se realiza un estudio para determinar las causas del porque los alumnos que realizan estudios universitarios los abandonan. Para lo cual construyen tres clasificadores donde observan que los resultados tienen sus particularidades y ventajas por esta razón no pueden ser comparados del todo. De acuerdo con el estudio son dos razones por las cuales explican este fenómeno. El cual de acuerdo a los autores es posible reducir las tasas de deserción de la universidad donde se realizó el estudio. Por otro lado, en (Torres, C. Z., Ramos, C. A., & Moraga, J. L. (2016)) realizan una investigación para clasificar los estudiantes desertores a partir de los datos socioeconómicos y académicos utilizando como metodología la CRISP-DM, para lo cual se analizan tres diferentes modelos de clasificación encontrando que Random Forest es el algoritmo con un mejor desempeño en general. De acuerdo con su estudio consideran que las variables académicas no resultaron significativas para la deserción de los alumnos.

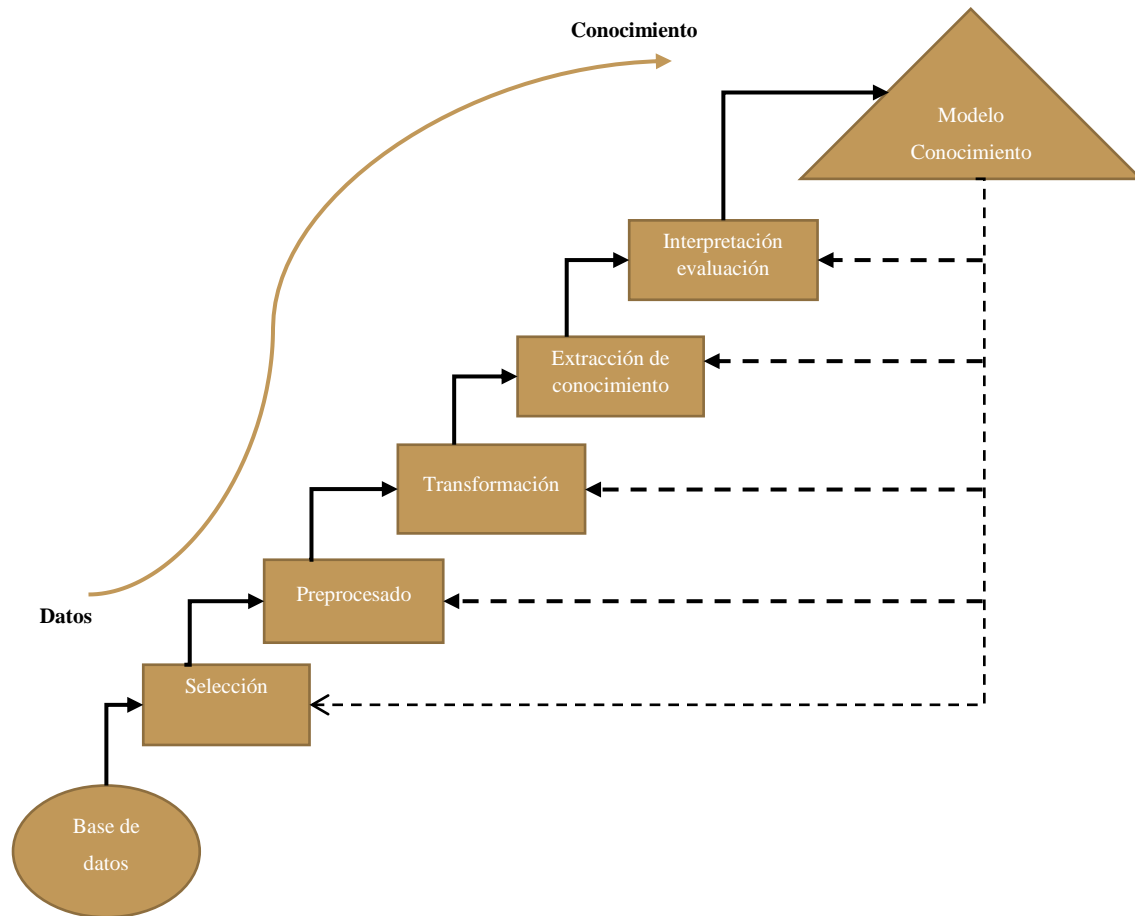
De igual forma en (Pérez-Gutiérrez, B. R. (2020)), se presenta un estudio de la comparación de diferentes técnicas para observar el fenómeno de deserción de los alumnos de una universidad de Colombia a partir del registro académico en un periodo de siete años considerando arboles de decisión, regresión logística y Naive Bayes y de esta manera establecer la mejor técnica que detecte los desertores. También utilizan Watson Analytics de IBM para comparar su usabilidad y precisión para un sistema no experto. Esperan que con su estudio puedan predecir las tasas de deserción escolar y de esta manera analizar y mejorar el rendimiento académico de los alumnos de la universidad.

2.1. Minería de datos

La minería de datos es una forma de explorar la información contenida en una base de datos mediante técnicas de aprendizaje supervisado, semisupervisado o no supervisado y obtener patrones, tendencias o reglas que ayudan a explicar el comportamiento de los datos en forma de conocimiento.

De acuerdo con (Witten & Frank, 2016) la minería de datos se puede definir como “el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos”. La tarea fundamental de la minería de datos es, encontrar modelos entendibles y legibles para el usuario a partir de un conjunto de datos.

El proceso que se sigue para obtener conocimiento de una base de datos mediante la aplicación de algoritmos de minería de datos se muestra en la figura 1, este proceso es conocido como KDD por sus siglas en inglés (Knowledge Discovery in Databases), se refiere al análisis de grandes bases de datos mediante diferentes algoritmos para obtener información útil para la organización (Tangarife Morales, 2016).

Figura 1 Proceso de extracción de conocimiento

A continuación se describe de forma general a que se refiere cada una de las etapas del Proceso de Extracción de Conocimiento (Hernández Orallo, Ramírez Quintana & Ferri Ramírez, 2010).

Etapa de selección

En la etapa de selección, una vez identificado el conocimiento relevante y prioritario y definidas las metas del proceso kdd, desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La selección de los datos varía de acuerdo con los objetivos del negocio.

Etapa de preprocesado

En la etapa de preprocesamiento/limpieza (data cleaning) se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos (missing y empty), datos nulos, datos duplicados y técnicas estadísticas para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista.

Etapa de transformación

En la etapa de transformación y/o reducción de datos, se buscan características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos.

Etapa de extracción de conocimiento

El objetivo de la etapa minería de datos es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación, clustering, patrones secuenciales y asociaciones, entre otras. Las técnicas de minería de datos crean modelos que son predictivos o descriptivos.

Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que se denominan variables objetivo, dependientes o clases, usando otras variables denominadas independientes o predictivas, como por ejemplo predecir para nuevos clientes si son buenos o malos basados en su estado civil, edad, género y profesión, o determinar para nuevos estudiantes si desertan o no en función de su zona de procedencia, facultad, estrato, género, edad y promedio de notas.

Etapa de interpretación/evaluación

En la etapa de interpretación/evaluación, se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. Por otra parte, se consolida el conocimiento descubierto para incorporarlo en otro sistema para posteriores acciones o, simplemente, para documentarlo y reportarlo a las partes interesadas; también para verificar y resolver conflictos potenciales con el conocimiento previamente descubierto.

2.1.2 Reglas de asociación

Las reglas de asociación son técnicas empleadas en el aprendizaje no supervisado para establecer posibles relaciones entre diferentes acciones aparentemente independientes entre sí, reconociendo como pueden ocurrir los eventos o acciones a partir de la aparición de otros. Así, este tipo de técnicas se usa cuando el análisis que se desea realizar es exploratorio (Pérez López & Santín González, 2008).

Existen diversos algoritmos para encontrar reglas de asociación como, A priori (véase Figura 2), Tertius, FPGrowth, FilteredAssociator, entre otros, el más utilizado es el algoritmo Apriori, el cual se basa en la búsqueda de los conjuntos de datos con determinada cobertura (Hernández Orallo, Ramírez Quintana & Ferri Ramírez, 2010). Este algoritmo es el empleado para este caso de estudio, ya que con él se generan las reglas a partir de conjuntos reducidos de datos que permiten una mejor formulación de estas, obteniendo así una mayor confianza y cobertura al ser evaluadas, a continuación, se describe el procedimiento a seguir para determinar las reglas.

1. Determinar la confianza y cobertura.
2. Se construyen los conjuntos formados por un solo item (conjunto) que superan la cobertura mínima.
3. Extraer del conjunto de reglas las que tengan un nivel de confianza mínimo.

Figura 2 Algoritmo A priori

ALGORITMO Apriori (D: datos, MinC: cobertura mínima)

```

1:   i=0
2:   Rellena_Item( $C_i$ )
3:   mientras  $C_i \neq 0$  hacer
4:       para X = elemento de  $C_i$  hacer
5:           si Cobertura(X)  $\geq$  MinC entonces
6:                $L_i = L_i * X$ 
7:           fin si
8:       fin para
9:        $C_{i+1} = \text{Selecciona\_Candidatos}(L_i)$ 
10:       $i = i + 1$ 
11:  fin mientras
12:  devolver C

```

2.2 Weka

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Contiene herramientas para la preparación de datos, clasificación, regresión, agrupamiento, minería de reglas de asociación y visualización. Weka es un software de código abierto emitido bajo la Licencia Pública General GNU. ("Weka 3 - Data Mining with Open Source Machine Learning Software in Java", 2021)

La plataforma Weka se caracteriza por tener:

- Disponibilidad: esta plataforma de software es libre gracias a la licencia pública general de GNU.
- Adaptabilidad: al estar implementada en lenguaje Java, es compatible casi con cualquier plataforma.
- Funcionalidad: está formada por un amplio repositorio de técnicas para preprocesamiento de datos y modelado.
- Sencillez: su uso es muy fácil gracias a su interfaz gráfica de usuario.

Al ejecutar el software brinda diferentes opciones para el procesado de los datos, los cuales son:

- Explorer: Es la interfaz visual de que permite trabajar fácilmente. Permite procesar, clasificar, asociar y visualizar datos intuitivamente sobre un sólo archivo de datos.
- Experimenter: Permite la comparación sistemática de una ejecución de los algoritmos predictivos de Weka sobre una colección de conjuntos de datos.
- KnowledgeFlow: Es la interfaz gráfica utilizada para desarrollar proyectos a través de flujos de información.
- Simple CLI: Simple Command-Line Interface, es una consola que permite acceder a todas las opciones de Weka desde línea de comandos.

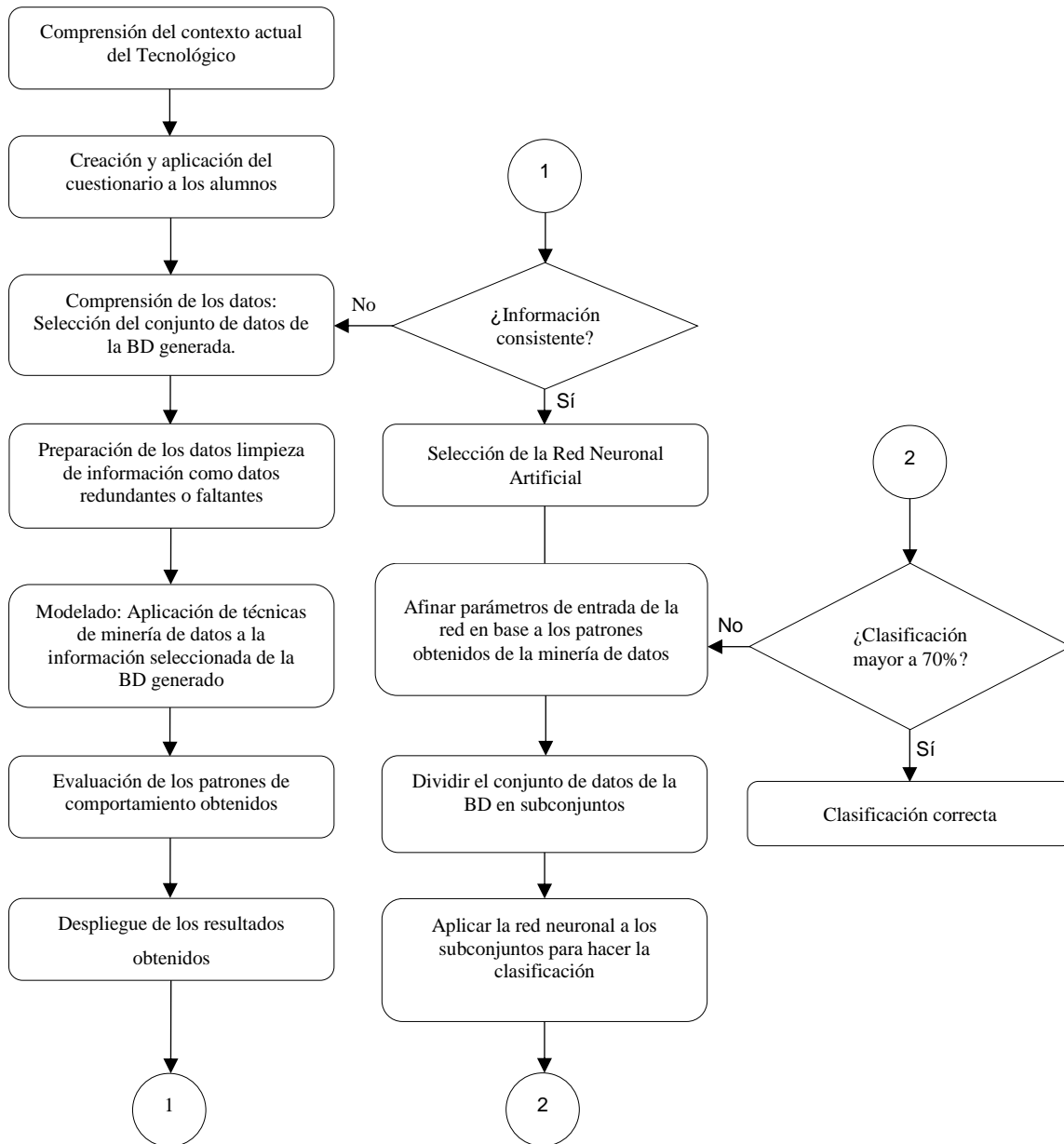
3. Metodología

El desarrollo del presente trabajo se lleva a cabo siguiendo la metodología presentada en la Figura 3, el diseño de ésta se basa siguiendo el proceso de extracción de características mostrado en la Figura 1.

En la Figura se observan 3 partes, donde la primera describe el proceso de minería de datos el cual consiste en la selección del conjunto de datos, pre-procesamiento, aplicación de algoritmos de minería de datos y su interpretación. La segunda parte de la metodología hace una evaluación de los resultados obtenidos de la primera etapa, si los resultados son consistentes se procede a la selección de la red neuronal, afinación de parámetros y se realiza la aplicación de la red neuronal. Finalmente, en la tercera parte se evalúa si la clasificación de la red neuronal fue mayor a 70% entonces la clasificación es correcta y con ello analiza el proceso.

Para poder llevar a cabo la experimentación se hace uso de una base de datos con un total de 75 atributos y 250 registros de alumnos que respondieron la encuesta para recabar los datos, es importante señalar que el número de muestras es pequeño, por esta razón en la primer fase de la metodología se hace el pre-procesamiento que consiste en generar variables a partir de los datos obtenidos, a condición de poder generar nuevos datos fue seguir el mismo comportamiento que tiene la base de datos original, esto se logró obteniendo la distribución normal de la base de datos, y en base a esta, se generaron los nuevos datos, dando un total de 1,000 registros.

Figura 3 Metodología a desarrollar



4. Resultados

La ejecución del algoritmo *A priori*, se realiza cuando la base de datos aún contiene los datos nominales, ya que éste algoritmo procesa la información nominal en lugar de datos numéricos como lo hacen la red neuronal y el algoritmo K-medias.

Una vez que se ha aplicado el algoritmo a la base de datos, se obtienen las reglas generadas, a partir de estas se eligen aquellas que tienen mayor representación, es decir, que la cobertura y precisión son mayores.

De las reglas obtenidas con la ejecución del algoritmo se realiza la siguiente interpretación, donde se muestra de forma sintetizada el comportamiento y factores que afectan el rendimiento de los alumnos y pueden ser detonantes de que determinado alumno deserte de la carrera. Dentro de las reglas de asociación que se encontraron en base a los resultados del algoritmo *A priori*, que tienen mayor relevancia en la base de datos en función de una mayor cobertura y precisión, se describen a continuación:

1. Estudiantes que no tienen hijos y han cursado en una institución pública la preparatoria no ha experimentado ninguna situación de penuria en los últimos seis meses.

2. Un alumno normal tiene en cuenta que el uso de una computadora para realizar sus actividades escolares es muy importante, tiene acceso a internet en casa, esto hace que sea mucho más fácil estudiar de forma independiente, además de considerar el entorno en el que vivir la universidad no es un límite para que abandone su educación.
3. De la misma forma que la regla anterior, un alumno regular es aquel que la carrera de sistemas informáticos fue su primera opción, conoce los contenidos de aprendizaje y se siente satisfecho con ellos, de la misma forma siente que falta el ambiente estudiantil. de un servicio de tutoría es un factor de abandono de los estudios.
4. La mayoría de los estudiantes regulares consideran neutral que una situación impactante pueda llevar a la deserción.
5. Aquellos estudiantes que viven solo con su madre y que no han experimentado ningún evento que haya ocurrido han alterado sus hábitos son alumnos regulares.
6. La mayoría de alumnos mencionan haber elegido la carrera de sistemas como su primera opción ya que es el área que siempre les ha gustado, de igual forma son alumnos que tienen internet en cada uno y no pasan por ningún evento que ha alterado sus hábitos normales en el último año. También se identificó a la mayoría de los estudiantes que tienen un promedio actual entre 80 y 89.
7. Es interesante saber que los estudiantes desconocen si existe un departamento encargado de la gestión de becas, aunque saben que tienen la posibilidad de tener una orientación con su tutor para que con su tutor puedan orientarlos ante cualquier situación o duda que los estudiantes puedan tener, resultando en un estudiante que puede abandonar fuertemente la competencia.
8. Al igual que la regla descrita anteriormente, incluso los estudiantes que de alguna manera abandonan la escuela muestran problemas cuando se trata de aprender.
9. Los estudiantes que no son desertores se caracterizan porque la carrera de sistemas informáticos pertenece al área de estudio que siempre les ha interesado, por eso mismo fue su primera opción cuando cursaban el nivel superior.
10. La carrera que estudian actualmente no fue su primera opción, esto podría significar que no es lo que en realidad quisieron estudiar o tal vez la escuela no fue su primera elección.
11. Su promedio actual está entre 80 a 89, este promedio podría ser considerado como regular, aun así, han presentado al menos un examen extraordinario.
12. Tienen una computadora personal pero su mayor forma de conectarse a internet es mediante datos celulares. En cuanto a los factores de apoyos económicos, no saben si el tecnológico cuenta con un departamento de becas, así como tampoco están muy bien informados sobre los apoyos que ellos pudieran recibir.
13. Están insatisfechos por alguno de los servicios que ofrece el Tecnológico, así como que consideran que algún factor por el cual pudieran pasar sería una razón para que abandonen sus estudios.

Conclusiones

En la presente investigación se desarrolló la primera fase del proyecto para la identificación de factores de riesgo en alumnos por la educación a distancia de la carrera de Ingeniería en Sistemas Computacionales del TESJo mediante la aplicación del algoritmo A priori. La información con la que se trabajó fue recolectada mediante aplicación de formularios a los estudiantes activos de la carrera de sistemas de los semestres primero, tercero, quinto, séptimo, noveno y décimo primero.

Todos los factores detectados fueron encontrados gracias a la aplicación de minería de datos a la base de datos, de manera que se puedan conocer los factores que provocan que un alumno deserte de la carrera por la situación que se vive actualmente.

Se considera que el número de muestras generado es relativamente corto dado que se manejan aproximadamente 1000 registros por lo que se considera que para la siguiente fase del análisis se lleven a cabo las siguientes acciones:

1. Aplicar los cuestionarios cada semestre e ir adaptándolos según las necesidades a situaciones cambio identificadas, esto permitirá tener constantemente datos nuevos y actuales con los cuales se puedan ir tratando con el paso del tiempo.
2. Aplicar algoritmos de redes neuronales para clasificar a los alumnos entre desertores y no desertores.
3. Aplicar otras técnicas de minería de datos, con el fin de ver que nuevas reglas o patrones podría predecir alguna otra técnica en cuestión.

Referencias

- Castrillón-Gómez, O. D., Sarache, W., & Ruiz-Herrera, S. (2020). Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formación universitaria*, 13(6), 217-228.
- Gómez, A. A. R., & Ibáñez, A. V. (2017). Aplicación de la Minería de Datos en la educación en línea. *Revista Colombiana de Tecnologías de Avanzada*, 1(29), 92-98.
- Hernández Orallo, J., Ramírez Quintana, M., & Ferri Ramírez, C. (2010). *Introducción a la minería de datos*. Madrid: Pearson.
- Panorama de la educación 2020, (2021). Retrieved 25 October 2021, from https://gpseducation.oecd.org/Content/EAGCountryNotes/EAG2020_CN_MEX_es.pdf
- Pérez-Gutiérrez, B. R. (2020). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. *Revista UIS Ingenierías*, 19(1), 193-204.
- Pérez López, C., & Santín González, D. (2008). *Minería de datos*. [Madrid]: Paraninfo Cengage Learning.
- Tangarife Morales, C. (2016). Estudio comparativo de metodologías para la detección de áreas de control de tensión. Colombia: Universidad Tecnológica de Pereira. Facultad de Ingenierías Eléctrica, Electrónica, Física y Ciencias de la Computación. *Ingeniería Eléctrica*.
- Terán, H. E. E., Saltos, M. A., de la Plata, C. M., & Terán, C. E. E. (2017). Implementación de minería de datos en la gestión académica de las instituciones de educación superior. *Didasc@ lia: Didáctica y Educación*, 8(3), 203-212.
- Witten, I., & Frank, (2016), E. *Data mining*. San Francisco, Calif.: Morgan Kaufmann.
- Garcés Fuenmayor, J., & Mora Bolaños, C. (2020). Estrategias de aprendizaje para mitigar la deserción estudiantil en el marco de la COVID-19. *Revista disciplinaria en ciencias económicas y sociales*, 23-45.
- Pérez Zuñiga, R., Martínez García, M., & Mena Hernández, E. (2021). Sars-CoV-2 en México y su efecto en los modelos educativos áulicos: una perspectiva crítica y reflexiva. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*.
- Rincón, I., Sir, S., & Suárez Castrillón, A. (2020). Impacto del programa Jóvenes en Acción en la deserción estudiantil en tiempos de Covid-19. *Espacios*, 296-303.
- Sandoval Elías, D. P. (2021). Recursos tecnológicos y conectividad: causas principales del ausentismo y deserción estudiantil. *Revista Docencia Universitaria*, 32-39.

Venegas Morales, M. C., & Rubio Hernández, M. (2021). Estudio exploratorio sobre educación remota durante COVID-19 en Universidad Tecnológica Santa Catarina. Efectos sociales, económicos, emocionales y de la salud ocasionados por la pandemia del COVID19.

Benavides-Lucksic, L., Villena-Presentación, R., & Andía-Valencia, W. (2020). La previsión de la deserción estudiantil en una Universidad Privada de Lima - Perú mediante procesos estocásticos, Cadenas de Markov. *Polo del Conocimiento*, 988-999.

Avila Pérez, M. L. (2021). Modelo de predicción de deserción estudiantil, apoyado en Tecnologías De Data Mining, en un curso de primera matrícula de la Escuela ECBTI De La UNAD. UNAD.

Benavides-Lucksic, L., Villena-Presentación, R., & Andía-Valencia, W. (2020). La previsión de la deserción estudiantil en una Universidad Privada de Lima - Perú mediante procesos estocásticos, Cadenas de Markov. *Polo del Conocimiento*, 988-999.

Contreras Bravo, L., Fuentes López, H. J., & González Guerrero, K. (2020). Transformación de la educación frente a la pandemia y la analítica de datos. *Revista Redipe*.

Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (2021). Consultado 28 October 2021, de <https://www.cs.waikato.ac.nz/ml/weka/>

Miranda, Mauricio A, & Guzmán, Jheser. (2017). Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos. *Formación universitaria*, 10(3), 61-68. <https://dx.doi.org/10.4067/S0718-50062017000300007>

Torres, C. Z., Ramos, C. A., & Moraga, J. L. (2016). Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos. *Ciencia Amazónica (Iquitos)*, 6(1), 73-84.

Pérez-Gutiérrez, B. R. (2020). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. *Revista UIS Ingenierías*, 19(1), 193-204.