

## Minería de datos aplicada para la identificación de factores de riesgo en alumnos

A. Reyes-Nava<sup>1,3</sup>, Allan Flores-Fuentes<sup>1</sup>, R. Alejo<sup>2</sup>, E. Rendón-Lara<sup>2</sup>

<sup>1</sup> Universidad Autónoma del Estado de México<sup>1</sup>,  
Estado de México, México

<sup>2</sup> Instituto Tecnológico de Toluca, Estado de México,  
México

<sup>3</sup> Tecnológico de Estudios Superiores de Jocotitlán,  
Estado de México, México

adriananava0@gmail.com, allan\_fflores@yahoo.com.mx, ralejoll@hotmail.com,  
erendonl@toluca.tecnm.mx

**Resumen.** En este trabajo se presenta la automatización del Sistema Integral de Tutoría del Tecnológico de Estudios Superiores de Jocotitlán para el área de Ingeniería en Sistemas Computacionales, el objetivo es la implementación del sistema en una plataforma digital y realizar un procesamiento en línea para que sea más eficiente el proceso de tutorías. El sistema consiste en la aplicación de cuestionarios y entrevistas a alumnos, obtenidos del manual del tutor del Sistema Nacional de Institutos Tecnológicos. La información obtenida es procesada en el software Weka aplicando técnicas de minería de datos, siguiendo la metodología Proceso de Extracción del Conocimiento con el propósito de facilitar la interpretación de resultados y encontrar conocimiento útil, acerca de los alumnos. Se analizan y discuten los resultados parciales del procesamiento a través de minería de datos de la información de generaciones del 2008 al 2014 de la carrera de Ingeniería en Sistemas Computacionales, encontrando relaciones entre aspectos familiares, económicos y académicos que indican un probable riesgo de deserción en alumnos con comportamientos comunes, estas relaciones sirven para proporcionar información suficiente y apropiada en la toma de decisiones en el tratamiento de la deserción escolar.

**Palabras clave:** proceso de extracción del conocimiento, tutoría, toma de decisiones, minería de datos.

### Data Mining Applied for the Identification of Risk Factors in Students

**Abstract.** In this paper presents the automation of the Tutoring System of the Technological Higher Education of Jocotitlán for the area of Engineering in

Computational Systems, the objective is the implementation of the system in a digital platform and perform online processing to be more efficient the process of tutorials. The system consists in the application of questionnaires and interviews to students, obtained from the tutor manual of the National System of Technological Institutes. The information obtained is processed in the Weka software by applying data mining techniques, following the Knowledge Extraction Process methodology in order to facilitate the interpretation of results and find useful knowledge about the students. We analyze and discuss the partial results of the processing through data mining of the information of generations from 2008 to 2014 of the career of Engineering in Computer Systems, finding relations between family, economic and academic aspects that indicate a probable risk of desertion in Students with common behaviors, these relationships serve to provide sufficient and appropriate information in decision-making in the treatment of dropout.

**Keywords:** knowledge extraction process, tutoring, decision making and data mining.

## 1. Introducción

La Tutoría es un proceso a través del cual se busca acompañar, asesorar, aconsejar y apoyar al estudiante para lograr en él una educación de calidad, integral y acorde a las necesidades de su entorno académico, laboral o personal [1]. En el año de 1998 la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) señaló la necesidad de modificar el proceso de enseñanza aprendizaje para dar respuesta a las necesidades del siglo XXI. Por su parte, en 2001 la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES) asume como eje rector para el mejoramiento de la educación la formación integrada, en la cual la tutoría tiene un papel distintivo. Asimismo, el plan de desarrollo Nacional 2013-2018 fija a la tutoría como medio para lograr disminuir la deserción escolar e incrementar los índices de eficiencia terminal. En ese mismo sentido, en los planes de desarrollo estatal e institucional del Tecnológico de Estudios Superiores de Jocotitlán se fija como meta el incremento de la eficiencia terminal.

El programa Institucional de Tutoría se implantó por primera vez en el Tecnológico de Estudios Superiores de Jocotitlán (TESJo) en el año 2008 y se operó por medio de los Jefes de División de cada carrera de acuerdo con los lineamientos establecidos por el Tecnológico Nacional de México (TECNM). En el 2010 se hizo un cambio, donde ya no eran los Jefes de División los encargados de la tutoría, sino los maestros. Para el 2012 se presentó el Manual del tutor, con la finalidad de que se convierta en una herramienta para el desarrollo de las actividades tutoriales.

A principios de 2014 el laboratorio de Minería de Datos y Reconocimiento de Patrones tomó como proyecto el estudio de la deserción escolar a través de técnicas de minería de datos. Para ello fue necesaria la construcción de una plataforma digital, en la cual, se pudiera desarrollar en línea el proceso de tutoría de acuerdo con lo establecido en el manual del tutor y adicionalmente se incorporaron algoritmos de minería de datos para el estudio de la deserción escolar. Actualmente, en el TESJo el

proceso de tutoría se rige con los lineamientos del Manual del Tutor del TECNM y es un proceso tradicional, en el sentido de que no se usan medios electrónicos o informáticos para las tareas de la tutoría. Sin embargo, en la carrera de Ingeniería en Sistemas Computacionales (ISC) se desarrolló un Sistema Integral de Tutoría del TESJo (SITTESJo), siendo la única carrera en llevar el proceso de tutoría de manera automatizada. Además se utilizan técnicas de minería de datos y reconocimiento de patrones para encontrar cuales son los factores que causan la deserción escolar.

## **2. Trabajo relacionado**

### **2.2. Sistema de tutorías**

El desarrollo de un sistema de tutorías en línea dentro del TESJo se realiza con la finalidad de hacer más eficiente el proceso de entrevista y recolección de información entre tutor y tutorado; de manera que al recabar la información de los alumnos pueda ser analizada para tener una perspectiva de quienes se encuentran en probable riesgo de deserción, fracaso estudiantil o inclusive saber quiénes son los alumnos destacados. Una vez que se obtienen los resultados de este análisis, la información es proporcionada a los tutores de cada grupo para emprender acciones tutoriales en su apoyo, con la finalidad de atacar a las problemáticas a las cuales se enfrenta la carrera de ISC como, la deserción escolar, inasistencias a clases, bajo rendimiento académico, entre otras.

Actualmente, en diferentes universidades se han desarrollado plataformas digitales para los procesos de tutoría. Por ejemplo, la Universidad Autónoma del Estado de México desarrolló el Sistema para la Tutoría Académica, en el cual se brinda el servicio de tutoría como apoyo al estudiante a través de un profesor del organismo académico que funge como tutor, se encarga de brindar el apoyo necesario y asesorar en los asuntos académicos como lección de unidades de aprendizaje, balance de créditos, canalización con asesores disciplinarios, entre otras actividades. Esto se lleva a cabo vía internet donde tutor y tutorado pueden interactuar para compartir información, analizar datos de diferentes tipos y formar un registro de las acciones emprendidas en su apoyo [2].

Por otro lado, [3] desarrolló un sistema de tutoría en línea para estudiantes de la carrera de ingeniería industrial de la Universidad del Valle de Atemajac, que surge por la necesidad de brindar una mejor educación a sus alumnos, tomando en cuenta los porcentajes de deserción y reprobación de las materias. Con la implementación del programa de tutoría se apoya a los alumnos desde el ingreso hasta el término de sus estudios haciendo uso del Programa General de Tutorías, asignando un tutor de apoyo desde el inicio de carrera y hasta el término de la misma.

También en [4], se hace una comparación entre los procesos de tutoría presencial y virtual en el Instituto Tecnológico de Sonora, donde los alumnos que están en la modalidad presencial, estadísticamente, tienen un promedio de calificaciones más bajo comparado con los que se encuentran en la modalidad virtual, lo cual hace notar que los alumnos prefieren un sistema de tutoría en línea, que un proceso tradicional.

No obstante, a pesar de que el problema de la tutoría se ha abordado de manera natural a través de herramientas digitales, la mayoría de los sistemas de tutoría son “tradicionales”, es decir, plataformas en las cuales se interactúa con el alumno a través

de un portal de internet, pero que no van más allá. En este trabajo se presenta al SITTESJo como una alternativa a los sistemas convencionales de tutoría. El SITTESJo tiene como característica principal la inclusión de algoritmos de minería de datos que permiten el análisis de los datos y su posible aprovechamiento para mejorar la relación entre tutelados y tutores, y de esta forma cumplir con uno de los principales retos de la tutoría, reducir la deserción escolar e incrementar los índices de eficiencia terminal.

## 2.2. Aplicaciones de la minería de datos para el análisis de la deserción escolar

Actualmente en las instituciones existen diversos problemas que afectan el desarrollo de la educación sin importar el nivel educativo que se curse. De manera específica uno de los principales problemas a los que se enfrenta el sistema educativo nacional es el abandono de los estudiantes en algún punto durante el trascurso de sus estudios [5]. Pueden existir diferentes factores por los cuales los alumnos deciden abandonar sus estudios antes de concluir y pueden depender de diferentes circunstancias, por ejemplo, el ambiente social, personal y económico del alumno.

La deserción escolar es un problema que existe en las instituciones educativas desde hace mucho tiempo. En [6] se describe a la deserción como el abandono de las labores de un estudiante, dejando de cumplir con sus obligaciones como estudiante inscrito en la institución, así como no asistir a clases de manera regular, lo cual afecta los índices de eficiencia terminal.

Hasta el día de hoy se han realizado diferentes estudios, no solo en México, sino también en diferentes países, donde se presenta el problema de la deserción escolar. El objetivo principal de estas investigaciones es conocer las causas que lo originan e identificar cómo se puede resolver o tratar.

En este contexto [7, 8] presentan investigaciones realizadas en las universidades nacionales de Argentina donde se busca explicar la causa de la deserción, haciendo una relación entre el rendimiento académico, estilos de aprendizaje y competencia espacial, esta se refiere a la capacidad de generar, recordar y transformar información simbólica mentalmente. Aplican metodologías descriptivas y estadísticas cualitativas para encontrar los factores que afectan el rendimiento académico y pueden ser causa de abandono de los estudios. Las conclusiones en estas investigaciones, muestran que se necesitan promover conductas académicas en los alumnos para que obtengan un aprendizaje autorregulado de manera que se favorezcan sus capacidades intelectuales. Por otro lado, se pudo determinar que los factores individuales como el ambiente social y cultural del alumno pueden afectar el desempeño académico.

Por ejemplo en [9] se hizo un estudio tomando como fuente interna de información la base de datos de control de admisiones y registro académico de la institución y como fuente externa se seleccionó información del Departamento Administrativo Nacional de Estadística del Sistema para la Prevención de la Deserción en la Educación Superior. De la información obtenida de estas bases de datos, se realizó el análisis mediante clustering para obtener los patrones de comportamiento comunes que son causa del abandono de los estudios, encontrando así que las principales causas son las bajas calificaciones en las asignaturas de los primeros semestres de la carrera, el pago de la colegiatura y la zona de procedencia del estudiante. A partir de los resultados que se

obtuvieron en esta investigación se pueden formular políticas y estrategias para incrementar los índices de retención estudiantil.

En [10] se hizo un análisis con información obtenida de estudiantes desertores, éste se realizó mediante proyecciones capaces de ordenar datos nominales y numéricos de manera supervisada, con el fin de encontrar los atributos más relevantes para ser utilizados en la aplicación de algoritmos de minería de datos, obteniéndose así resultados claros que permitan la toma de decisiones en favor de la disminución de la deserción.

Con base en lo anterior se puede concluir, que el uso de tecnologías emergentes como la minería de datos para identificar a los alumnos desertores o con problemas académicos en las instituciones, ha sido tema de estudio para diferentes instituciones educativas.

### **3. Elementos que constituyen el proceso de tutoría del SITTESJo**

El SITTESJo consiste de una serie de cuestionarios que el alumno debe contestar en forma manual, los cuales son establecidos en el manual del tutor del Sistema Nacional de Institutos Tecnológicos (SNIT) para el proceso de tutoría. De manera interna el TESJo asigna a algunos de ellos códigos específicos (FO-TESJo-039, FO-TESJo-040), mientras que a otros no los considera en el proceso formal de tutoría. No obstante, el SITTESJo incluye todos los cuestionarios y pruebas propuestas en el Manual del Tutor del SNIT, e inclusive incluye un módulo de calificaciones. A continuación, se describen brevemente los cuestionarios y pruebas incluidas en el SITTESJo.

1. FO-TESJo-039: Aquí se incluyen preguntas sobre aspectos familiares, socioeconómicos y académicos de los alumnos, a manera de conocer sus antecedentes; esta información se almacena en la base de datos del sistema y puede ser modificada en cada aplicación de los cuestionarios si es que existe algún cambio en la situación del alumno [11].
2. FO-TESJo-040: Este cuestionario incluye preguntas para conocer de manera específica si el alumno trabaja y en qué, quién y cómo solventa sus gastos, a quién mantiene informado de sus estudios y su estado de salud de manera general [11].
3. Test de autoestima, de asertividad, de estilos de aprendizaje y de habilidades de estudio, éstos son cuestionarios de preguntas cerradas para reconocer problemas referentes a organización de actividades, técnicas para sus estudios y su motivación a estudiar [11].
4. Calificaciones. De manera opcional se agregó el apartado de calificaciones donde los alumnos capturan la calificación obtenida por cada unidad de aprendizaje de los diferentes programas educativos que cursan.

#### 4. Minería de datos y algunos clasificadores

Las bases de datos que existen actualmente contienen una gran cantidad de información que en muchas ocasiones no es aprovechada, porque no existe suficiente personal que se encargue de realizar el análisis de los datos. No obstante, en las últimas décadas se han popularizado nuevas formas de analizar estos datos y obtener información en forma de conocimiento; por medio del empleo de herramientas computacionales, es decir, a través del uso de la minería de datos. De acuerdo con [12] 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.

Otros autores como [13] definen de manera simple a la minería de datos como “el proceso de convertir datos en conocimiento”. Esta definición es la que se puede comprender de manera más sencilla y enfoca el punto central de la minería de datos.

Para poder encontrar conocimiento a partir de un conjunto de datos es necesario tomar en cuenta el proceso de descubrimiento de conocimientos en bases de datos, en este caso se aplica KDD (Knowledge Discovery in Databases). Este es un proceso enfocado a identificar patrones válidos, potencialmente útiles y comprensibles a partir de un conjunto de datos, este proceso incluye no solo la obtención de patrones, sino también se realiza la interpretación de los resultados de manera que apoye a la toma de decisiones [14].

El proceso de descubrimiento de conocimiento en bases de datos consta de diferentes etapas, en la Figura 1 se muestra de manera esquemática cómo se lleva a cabo este proceso. Se parte desde un sistema de información de donde se obtienen los datos a ser analizados, los cuales pasaran por el proceso de preparación, aplicación de técnicas y algoritmos de minería de datos, obtención de patrones y evaluación o interpretación de los mismos para llegar a la obtención del conocimiento que ayuda a la toma de decisiones en el proceso de negocios.

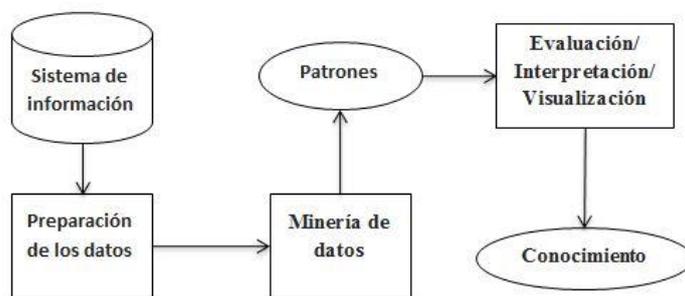


Fig. 1. Proceso KDD.

En el proceso de descubrimiento de conocimiento se identifica la etapa de minería de datos, donde se eligen los algoritmos a utilizar de acuerdo con la información con la que se cuenta y al conocimiento que se quiera obtener.

Un ejemplo de ello son los métodos bayesianos, los cuales se centran en el estudio descriptivo y predictivo, porque los algoritmos se enfocan en el descubrimiento de relaciones de independencia y relevancia entre las variables. Los métodos bayesianos se usan para realizar inferencias a partir de los datos y a partir de modelos probabilísticos usados para formular hipótesis [15].

Otro ejemplo son las redes neuronales artificiales, estas pueden ser usadas en aprendizaje supervisado o no supervisado. Una red neuronal artificial es en sí un método de aprendizaje cuya finalidad es la de simular procesos biológicos de información, las cuales parten de la capacidad del ser humano para procesar información [13].

#### **4.4. 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 [16].

Existen diversos algoritmos para encontrar reglas de asociación, el más utilizado es el algoritmo Apriori, el cual “se basa en la búsqueda de los conjuntos de datos con determinada cobertura” [13]. Este algoritmo es el empleado para este caso de estudio, a continuación se presenta el pseudocódigo para determinar las reglas.

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

A partir de lo anterior, el algoritmo queda de la siguiente manera:

```
ALGORITMO Apriori (D: datos, MinC: cobertura mínima)
  i=0
  Rellena_Item(Ci)
  MIENTRAS Ci != 0
    PARA_CADA X=elemento de Ci
      SI Cobertura(X) >= MinC ENTONCES
        Li=Li * X
    FIN_PARA
  Ci+1=Selecciona_Candidatos(Li)
  i=i+1
  FIN_MIENTRAS
  RETORNA C
FIN_ALGORITMO
```

## 4.2. Clustering

El clustering o agrupamiento es una técnica de minería de datos, la cual consiste en formar grupos, haciendo una división de los datos en grupos de objetos similares con base en un criterio de cercanía, aumentando la similitud de los elementos de un mismo grupo. Uno de los algoritmos de clustering o agrupamiento más común es el propuesto por MacQueen, llamado Simple K-Means, donde K determina el número de clusters o grupos que se desean obtener [17].

El algoritmo K-means inicia seleccionando o calculando los centroides iniciales, dependiendo del criterio de selección de centroides, posteriormente asigna los objetos a su centroide más cercano, para después recalcular los nuevos centroides esto lo realiza hasta que el algoritmo converge. El pseudocódigo se presenta a continuación:

Paso 1. Selecciona los K centroides iniciales.

Paso 2. Asigna los objetos  $X_i$  del conjunto de datos X, a su centroide más cercano.

Paso 3. Re-calcula los nuevos centros, regresa al paso 2, hasta que el algoritmo converge.

## 5. Metodología para el análisis de los datos almacenados por el SITTESJo

El análisis de los datos fue realizado a través de dos importantes herramientas de minería de datos: clustering y reglas de asociación; en particular, se utilizó el algoritmo de agrupamiento K-means y el algoritmo Apriori para las reglas de asociación. Para llevar a cabo el proceso de análisis de los datos se empleó la metodología KDD, la cual dentro de este trabajo consistió en:

1. Seleccionar los datos y análisis de sus propiedades. Una vez aplicados los cuestionarios a los alumnos y que la información está en la base de datos, se determina cuáles son los datos relevantes para hacer el análisis de la información disponible como antecedentes académicos, familiares y socio-económicos.

En este paso para reglas de asociación de utilizaron datos específicos de la base de datos, los cuales fueron: práctica de deporte, problemas económicos, si trabajan o no, promedio de preparatoria, interrupciones en sus estudios y el campo de acción de la carrera.

Para el caso de agrupamiento se emplearon los datos del porqué eligió la carrera, tratamientos médicos, dependencias económicas (si están casados o tienen hijos) y conocimiento de programas como la becas.

2. Pre-procesamiento de datos. En esta etapa se hace la limpieza de la base de datos, encontrando los datos inconsistentes, fuera de rango, faltantes o campos vacíos para ser integrados posteriormente y puedan ser utilizados en el análisis con

minería de datos, en esta etapa se convierte el archivo en formato .arff, el cual es el aceptado para el análisis con weka, este es un software diseñado para el análisis de bases de datos, aplicando diferentes algoritmos tanto de aprendizaje supervisado como no supervisado para obtener estadísticas y patrones de tendencia según sea lo que se estudie.

3. Aplicación de técnicas de minería de datos. Una vez que se obtienen los archivos en formato .arff, se procede a su uso en la herramienta weka, donde se carga el archivo y se procede a elegir la opción por la cual será procesada la información, para el caso de reglas de asociación, se elige la opción *associate* y posteriormente se selecciona el algoritmo *apriori* y se ajustan parámetros como el número de reglas, la salida, la cobertura mínima, entre otras, para posteriormente iniciar la ejecución y obtener los patrones.

En el caso de clustering, se sigue un procedimiento similar, seleccionando la opción *cluster*, después el algoritmo a emplear, en este caso *k-means* y se ajustan los parámetros, como el número de clusters en los que se va a agrupar la información y las iteraciones que tendrá.

4. Interpretación y evaluación. A partir de los resultados obtenidos de la fase anterior, se analizan los “patrones” obtenidos por weka y se evalúa si realmente son útiles, esta tarea es realizada por el analista.

## **6. Resultados**

Esta sección presenta los principales resultados obtenidos del análisis de los datos del SITTESJO a través de técnicas de minería de datos. Se divide en dos subsecciones, una para la discusión de resultados obtenidos con reglas de asociación y otra para los resultados obtenidos del análisis por medio de herramientas de agrupamiento.

### **6.1. Reglas de asociación**

Para el análisis mediante reglas de asociación se utilizó la información de la base de datos que corresponde a los apartados de antecedentes económicos, familiares y académicos, a partir de esta información se llegaron a las siguientes conclusiones.

1. Los alumnos que suelen practicar algún tipo de deporte como el fútbol a su vez tienen problemas económicos y los que no practican ninguna actividad deportiva no presentan problemas económicos.
2. La mayor parte de los alumnos que deserta tienen problemas económicos, razón por la cual deben buscar una forma de obtener ingresos y deciden entrar a trabajar, y al pasar el tiempo deciden abandonar sus estudios para continuar obteniendo el ingreso económico.
3. Los alumnos que desertan son alumnos que en sus estudios previos han sido buenos con un promedio favorable, lo cual indica que la causa de su deserción no se debe directamente a situaciones académicas.

4. La mayoría de los alumnos que desertan nunca han interrumpido sus estudios en los niveles básico y medio superior.
5. La gran mayoría de los alumnos que ingresan a la carrera de ingeniería en sistemas computacionales tienen una idea errónea del enfoque de esta al momento en el que deciden inscribirse.

A partir de los resultados obtenidos con el uso de reglas de asociación se obtuvo un panorama general de la situación de los alumnos de la comunidad estudiantil de la carrera de ISC y las causas probables por las que deciden abandonar sus estudios. Por ejemplo, la contraposición de opiniones entre alumnos (hombres y mujeres) y la relación que existe entre estudiantes que practican deportes y su situación económica, da una referencia de como los alumnos pueden actuar y de lo que piensan. Estos factores pueden influir en que los alumnos tengan un bajo desempeño académico y pueda ser esto un detonante en decisiones como abandonar sus estudios. Sin embargo, estos resultados son preliminares y es necesario un estudio más extenso para poder obtener más información que ayude a la toma de decisiones por parte de la comunidad académica. Los resultados presentados corresponden a las generaciones del 2008 al 2013.

## 6.2. Clustering o agrupamiento

Una vez concluida la aplicación de reglas de asociación se utilizó la técnica de clustering k-medias, a un conjunto de datos más amplio, ya que se incluyeron nuevos datos obtenidos de la aplicación del cuestionario FO-TESJo-039 y los test de autoestima, asertividad y estilos de aprendizaje, además del registro de nuevas calificaciones de los alumnos. De la aplicación de este algoritmo se obtuvieron los siguientes resultados:

1. De un total de 831 alumnos, 269 de ellos tienen otra licenciatura en mente que les hubiera gustado estudiar, de estos 269, 15 prefieren no seguir estudiando y comenzar a trabajar para obtener un ingreso económico y el resto prefiere salir de la carrera para buscar el ingreso a la carrera de su agrado.
2. De un 60% de alumnos que toman medicamentos, sus padres tienen un nivel educativo superior al de secundaria por lo que cuentan con atención médica, pero se debe dar seguimiento para que no sea causa de deserción. En este caso la atención es porque del 100% de alumnos que toman medicamentos solo el 60% cuenta con atención médica mientras que el otro 40% no cuenta con la atención, lo cual, en algunos casos no es suficiente o bien sus padres no los apoyan al no saber del problema.
3. Los alumnos que tienen a alguien que depende económicamente de ellos, es decir, son casados o que trabajan son fuertes candidatos a desertar debido a su situación económica y familiar.
4. Todos los alumnos encuestados mencionan que conocen el campo de acción de la carrera y el 91% de ellos no ha estado inscrito en otra licenciatura, sin embargo únicamente el 4% de los alumnos no tiene motivos para no concluir la carrera, esto es que la carrera fue su primera opción y no tienen otra licenciatura en mente,

de manera que las causas de deserción se deben a causas sociales, económicas, personales o bien que no tienen el suficiente compromiso para culminar sus estudios.

5. A los alumnos que tienen otra carrera en mente se les dificulta la situación escolar ya que están en clases solo porque no tienen otra opción por el momento pero llegada su oportunidad ellos se decidirán por abandonar la carrera e inscribirse para tratar de entrar a lo que ellos realmente quieren estudiar.
6. La situación económica del estudiante propicia deserción cuando no se dan a conocer cuáles son las becas a las cuales el alumno puede acceder ya sean por parte de la institución o bien por algún organismo gubernamental, esta es tarea de los tutores o bien del personal de servicios escolares de la institución, pero en muchas ocasiones se deja de lado y no se presta la atención suficiente a la publicación de las convocatorias entre los alumnos.
7. De los alumnos enfermos se observa que el seguro institucional ayuda a cubrir las necesidades del caso, ya que de no tenerlo no podrían cubrir sus necesidades, lo cual lleva a que el alumno en ocasiones este estudiando por conservar el seguro y no porque realmente quiera hacerlo, lo cual se ve reflejado en las calificaciones de este al término de cada semestre.

Con los resultados obtenidos de este análisis, se puede observar que conforme se obtienen más datos de la aplicación de los cuestionarios, el análisis se hace más profundo y se obtiene información más precisa sobre las necesidades de los alumnos, lo cual representa un beneficio en la toma de decisiones que ayuden a la permanencia de los alumnos en la institución y puedan concluir la carrera de manera satisfactoria.

## **7. Conclusiones y trabajo futuro**

En este trabajo se presenta el desarrollo del SITTESJo, enfocado a la automatización del proceso de tutoría para hacer este proceso más eficiente y eficaz, y que a la vez permita la adecuada toma de decisiones para atender problemas fundamentales como la deserción escolar.

El trabajo que se ha realizado hasta el momento sobre el desarrollo del SITTESJo se enfoca en la captura de información de los alumnos de ISC misma que es almacenada en la base de datos, para posteriormente ser analizada mediante técnicas de minería de datos y obtener patrones de comportamiento comunes entre los alumnos, con los cuales se pueden identificar problemas presentes en la comunidad estudiantil que afectan su desempeño y ocasionan problemas tanto para el alumno ya que puede no concluir sus estudios, como para la institución misma en el incumplimiento de la eficiencia terminal. La principal aportación del trabajo discutido en este artículo, es que además de sistematizar el proceso de tutoría, apoya el análisis de los datos por medio de herramientas de minería de datos y reconocimiento de patrones.

Con el análisis de la información que se realizó se obtuvieron resultados preliminares de los factores que pueden influir en la deserción y bajo desempeño de los estudiantes de la carrera de ISC del TESJo, entre estos se encuentran factores sociales, económicos y académicos.

Finalmente mediante este proyecto se el desarrollo un sistema integral de tutorías, que consiste en la página web y la aplicación de técnicas de minería de datos para obtener patrones. Sin embargo, como trabajo a futuro, es necesario probar con otras técnicas como un sistema experto basado en redes neuronales artificiales, para mejorar los resultados obtenidos hasta el momento.

## Referencias

1. García-Ibarra, C. A.: Manual del tutor del SNEST. México, SEP, pp. 145 (2012)
2. Sistema Inteligente para la Tutoría Académica. Disponible en: [https://www.sita.uaemex.mx/tutoria/index\\_ok3.html](https://www.sita.uaemex.mx/tutoria/index_ok3.html), México, SEP (2015)
3. Dávila-Avenida, M.: Implementación de tutoría en línea para alumnos de la facultad de ingenierías de la universidad del valle de Atemajac en un entorno virtual de aprendizaje. En: X Congreso Nacional de Investigación Educativa, pp. 1–13 (2008)
4. García-López, R., Cuevas-Salazar, O., Vales-García, J., Cruz-Medina, I.: Impacto de la tutoría presencial y virtual en el desempeño académico de alumnos universitarios. Revista Iberoamericana De Educación, Vol. 58, No. 2, pp. 1–11 (2015)
5. Secretaría de Educación Pública: Programa sectorial de educación 2013-2018, México, SEP, pp. 117 (2013)
6. Chaín, R. R.: Deserción, rezago y eficiencia terminal en las IES: Propuestas metodológicas para su estudio. México, ANUIES, 1 Ed, pp. 244 (2001)
7. García de Fanelli, A.: Rendimiento académico y abandono universitario: Modelos, resultados y alcances de la producción académica en la Argentina. Revista Argentina de Educación Superior, Vol. 8, pp. 9–38 (2014)
8. Maris-Vázquez, S., Noriega-Biggio, M., Maris-García, S.: Relaciones entre rendimiento académico, competencia espacial, estilos de aprendizaje y deserción. Revista Electrónica de Investigación Educativa, Vol. 15, No. 1, pp. 29–44 (2013)
9. Tamirán-Pereira, R., Calderón-Romero, A., Jiménez-Toledo, J.: Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. Revista Vínculos, Vol. 10, No. 1, pp. 373–383 (2013)
10. Formia, S., Lanzarini, L., Hasperué, W.: Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio. Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación, Vol. 11, pp. 92–98 (2013)
11. Secretaría de Educación Pública: Formatos codificados TESJo. Disponible en: [http://tesjo.edomex.gob.mx/formatos\\_codificados](http://tesjo.edomex.gob.mx/formatos_codificados), México, SEP (2016)
12. Witten, I., Frank, E.: Data mining. San Francisco, Calif., Morgan Kaufmann, 4 Ed., pp. 664 (2016)
13. Hernández-Orallo, J., Ramírez-Quintana, M., Ferri-Ramírez, C.: Introducción a la minería de datos. Madrid, Pearson Prentice Hall, 1 Ed., pp. 680 (2004)
14. Valcárcel-Asencios, V.: Data Mining y el descubrimiento del conocimiento. Revista de la Facultad de Ingeniería Industrial, Vol. 7, No. 2, pp. 83–86 (2004)
15. Mesa-Páez, L. O., Rivera-Lozano, M., Romero-Dávila, J. A.: Descripción general de la Inferencia Bayesiana y sus aplicaciones en los procesos de gestión. La Simul. al Serv. la Acad. Fac. Adm. Univ. del Rosario, Vol. 2, pp. 1–28 (2011)
16. Pérez-López, C., Santín-González, D.: Minería de datos: Técnicas y herramientas. Madrid, España, Thomson, 2 Ed., pp. 808 (2007)

17. MacQueen, J. B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Symposium on Math, Statistics, and Probability, Berkeley, CA, University of California Press, pp. 281–297 (1967)