

# Estrategia de minería de datos para la identificación de jóvenes en riesgo de abandono escolar en preparatorias

Erick Alonso Castro-Navarro, José Luis Ochoa-Hernandez, Gerardo Sánchez-Schmitz

Universidad de Sonora, Departamento de Ingeniería Industrial, Rosales y Blvd. Luis Encinas S/N CP. 83000, Hermosillo, Sonora, México.  
erick.cn89@gmail.com, joseluis.ochoa@industrial.uson.mx,  
gsanchez@industrial.uson.mx

**Resumen.** La retención de los alumnos en las instituciones de educación media superior es de gran importancia para un país, ya que mientras más preparado se encuentren, mejor futuro tendrá él y beneficiará a los involucrados. En la presente investigación se estudia el problema de deserción que tienen los alumnos de preparatoria, enfocándose en el número de alumnos que desertan, tratando de reducir al máximo estos casos. Actualmente, la preparatoria cuenta con un área de orientación educativa, la cual intenta atacar, estudiar y resolver este problema, sin embargo, el proceso que se lleva a cabo para su identificación no está dando los resultados esperados, además de que emplea más de la mitad del tiempo del semestre. Para ello, se propone implementar un estudio basado en Minería de Datos, que ayude a identificar de forma eficiente aquellos alumnos que presenten riesgo de abandonar sus estudios y puedan ser atendidos de forma temprana.

**Palabras clave:** minería de datos, deserción estudiantil, abandono escolar.

## 1 Introducción.

Las tecnologías de la información y la comunicación nos han proporcionado una nueva visión del mundo que conocemos, reduciendo el tiempo dedicado a las tareas que comúnmente podrían llevarnos días, semanas, meses e incluso años [7].

También han hecho posible que la información sea más fácil de capturar, almacenar y distribuir [1], de igual manera han hecho que esta sea más difícil de analizar, procesar, comprender e interpretar. Las organizaciones, por lo tanto, requieren comprender toda esa información para que les ayude a explotar de manera eficaz esa fuente de datos [4].

La Minería de Datos (MD) es un concepto que ha trascendido como un ejemplo de avance tecnológico, la cual proporciona a las organizaciones la oportunidad de realizar análisis muy complejos en menor tiempo y con mayor eficacia [11], pudiéndose aplicar en diferentes sectores, tanto científicos, ingenieriles o comerciales. Una de las finalidades de

Castro-Navarro EA, Ochoa-Hernandez JL, Sánchez-Schmitz G (2018) Estrategia de minería de datos para la identificación de jóvenes en riesgo de abandono escolar en preparatorias. Avances de Investigación en Ingeniería en el Estado de Sonora 4 (1):208-213

la Minería de Datos, es la de encontrar patrones que nos acerquen al comportamiento de los datos para poder comprenderlos.

Este concepto, se ha introducido en el nivel educativo para dar solución a diversas situaciones. Una de ellas, es la predicción, tratando de etiquetar a aquellos jóvenes que se encuentra en riesgo de abandonar la escuela y/o identificar aquellos motivos que dieron origen a este suceso.

Las instituciones educativas de nivel Medio Superior, buscan ofrecer una sólida formación integral y de calidad a sus alumnos, que les permita acceder a sus estudios superiores o introducirse al ámbito laboral para la realización de sus proyectos de vida. La Secretaría de Educación Pública define la deserción escolar como el abandono de las actividades escolares antes de terminar algún grado o nivel educativo, se conoce que la eficiencia terminal es insatisfactoria, ya que solo el 57% termina el bachillerato.

Es por ello por lo que se pone en operación el departamento de orientación educativa por la Dirección General de Bachillerato (DGB), con el objetivo de prevenir, formar e integrar al alumno del bachiller, con actividades que le permiten detectar y atender los factores de riesgo para evitar la deserción escolar.

El objetivo de esta investigación es realizar una propuesta de Minería de Datos que mejore el proceso de identificación de estudiantes con riesgo de abandonar sus estudios, empleando los instrumentos actuales y tratando de hacerlo en el menor tiempo posible, ya que el tiempo actual ocupa prácticamente la mitad del semestre.

El artículo se ha estructurado de la siguiente forma: En el punto 2 se presenta el marco teórico que muestra una revisión de la literatura y estudios previos que brindan el conocimiento sobre las herramientas y estrategias que están siendo utilizadas en diferentes partes del mundo en el sector educativo. Enseguida viene el punto 3 con la descripción del problema, el punto 4 que describe la metodología que se intenta seguir. El punto 5 presenta los resultados y el 6 las conclusiones.

## **2 Marco Teórico.**

Las tecnologías de la información y la comunicación (TIC) se han convertido en una de las herramientas más significativas para cualquier organización que desee recopilar y analizar de forma eficaz y eficiente los datos generados día a día [13]. Las aplicaciones tecnológicas y la implementación de técnicas de inteligencia artificial están destinadas ayudar a tomar decisiones en escenarios donde la cantidad de información complica el análisis por los expertos [4], por tanto, para realizar un análisis a profundidad de forma automática, han surgido una serie de técnicas que facilitan el procesamiento avanzado de los datos, sin embargo, es la transformación a conocimiento y la aplicación de éste, lo que genera valor [2].

Entre las tecnologías más relevantes en los últimos años para analizar abundantes cantidades de información se encuentra la minería de datos (MD), la cual se ha vuelto indispensable para examinar y obtener resultados [5]. La MD es una nueva tecnología de manejo y análisis de información que aprovecha la capacidad existente no solo para

procesar, almacenar y transmitir datos a gran velocidad y a bajo costo, sino también para hallar un contenido específico dentro la diversidad de fuentes existentes, lo cual resulta muy útil para las organizaciones, pues les permite tomar decisiones mejor fundamentadas para su futuro [1]. El proceso de extracción del conocimiento contempla siete pasos los cuales son: selección, procesamiento, transformación, MD, interpretación y evaluación. Donde su meta es utilizar grandes volúmenes de datos para descubrir patrones ocultos y relaciones útiles en la toma de decisiones en base a modelos predictivos y descriptivos [15].

El análisis predictivo es un proceso que permite estimar eventos a raíz del análisis de datos históricos y presentes a través de métodos estadísticos y reconocimiento de patrones [8]. Por esto, existe una tendencia al uso de la MD en el área de educación [10], la minería de datos en la educación (MDE) no es un concepto nuevo, su estudio y aplicación ha tomado mayor relevancia en los últimos años, ya que permite encontrar patrones que caracterizan el comportamiento de los estudiantes con base en sus logros, evaluaciones y dominio del contenido de conocimiento [3].

En un estudio realizado por [13] debido a la preocupación por los altos índices de abandono escolar en México, elaboró un diseño para la automatización de un modelo predictivo del rendimiento académico de los estudiantes en el Politécnico Nacional, ya que se contaba con un bajo rendimiento académico de los estudiantes. Donde se obtuvo un 73% de exactitud en la predicción, utilizando el proceso completo de la aplicación de las técnicas de MD basándose en el algoritmo de Naïve Bayes.

En un estudio realizado por [12], se realizó un sistema para la automatización del área de tutorías de la Universidad de Jocotitlán, Estado de México, debido a que se contaba con alto grado de abandono escolar, el departamento decide aplicar estrategias de MD para mejorar la toma de decisiones por parte de la institución. Este sistema se basa en la aplicación de encuestas de forma digital, las cuales son analizadas por la herramienta de minería de datos WEKA, donde se logra obtener patrones de comportamientos comunes entre los estudiantes y resultados preliminares de los factores que influyen en la deserción y el bajo desempeño de los estudiantes.

### **3 Descripción del problema abordar.**

El Colegio de Bachilleres del Estado de Sonora, cuenta con el área de orientación educativa (OE), que tiene como objetivo detectar a los alumnos que tengan alguna probabilidad de abandonar la institución, sin embargo, el área tiene dificultades para realizar la detección adecuada de los estudiantes, lo que genera que los alumnos con problemas académicos y psicológicos no puedan ser detectados oportunamente, ocasionando que estos abandonen la institución.

Actualmente el proceso de detección de alumnos en riesgo se basa en un cuestionario que se aplica al inicio de cada semestre, este cuestionario contiene 49 preguntas divididas en 7 aspectos: 1) Datos generales, 2) situación laboral, 3) Situación cultural, 4) condiciones de estudio, 5) valoración familiar, 6) condiciones de salud y 7) hábitos alimenticios y rendimiento académico. De las 49 preguntas, los docentes consideran como riesgo, aquel

joven que responde positivamente a por lo menos una de las siguientes preguntas; ¿Piensas que no cuentas con las habilidades adecuadas para salir adelante?, ¿Has estado a estado a punto de quitarte la vida?, ¿Has pensado que vale más morir que vivir?, esto debido a que la información es analizada de forma manual, esto provoca que se omita toda la información correspondiente de los demás rubros de la encuesta, lo que indica que existe una gran cantidad de información que no es procesada para su análisis y toma de decisiones.

Una vez aplicado el cuestionario, la información es examinada, en base a sus respuestas, los orientadores determinan los casos positivos y se entrevistan con ellos para corroborar la información que el alumno generó en el cuestionario, en base a esa percepción, se toman las decisiones. Actualmente este proceso está llevando un tiempo estimado de 45 días, lo que representa aproximadamente un 57% de avance del semestre, es decir, se ha cursado más de la mitad de este.

Esto denota que el proceso de detección realizado por los orientadores no se lleva a cabo de forma eficaz y eficiente, lo que provoca que gran parte de los alumnos no sean detectados a tiempo y abandonen la institución.

#### 4 Desarrollo de la solución.



Figura 19. Propuesta de solución (Desarrollo personal)

La propuesta de solución consta de 4 etapas principales:

1) Recopilación de información: se deberá elaborar, retomar o modificar un instrumento para la recolección de datos, donde se deberán considerar los factores principales que implican que un alumno deserte. 2) Transformación de datos: los datos obtenidos deberán seguir un reprocesamiento, es decir, seleccionar la información que es realmente relevante,

además de estandarizar aquellos donde las preguntas sean abiertas para que la minería de datos sea más eficiente. 3) Minería de datos: Deberán aplicarse las técnicas de minería de datos para obtener información acerca del comportamiento de los datos con la finalidad de extraer patrones que no ayuden a identificar jóvenes en riesgo. 4) Evaluación: Deberá realizarse una evaluación, para determinar la efectividad de las herramientas de minería de datos, esta evaluación se realizará en función a la comparación de los resultados detectados por el área de orientación educativa y de la herramienta de minería de datos.

## 5 Resultados.

Con el apoyo de las tecnologías de la información y la comunicación se busca establecer un nuevo proceso basado en Minería de Datos que mejore el método de análisis y los tiempos de detección de los alumnos en riesgo. Permitiendo que los orientadores centrarse en aquellos alumnos que están en riesgo de una manera oportuna.

## 6 Conclusiones.

En la revisión de los casos de estudio se puede observar investigaciones similares a lo propuesto en esta investigación, por lo que podemos mencionar que la aplicación de herramientas de minería de datos nos apoya en la mejora la detección y el análisis de jóvenes en riesgo, por tanto, se espera implementar algunas herramientas de minería de datos con la finalidad de reducir los tiempos de detección realizados e incrementar la eficacia en el departamento de manera significativa.

## Referencias

1. Altamiranda, L., Peña, A. M., Ospino, M., Volpe, I., Ortega, D. y Cantillo, E., 2013. Minería de datos como herramienta para el desarrollo de estrategias de mercadeo B2B en sectores productivos, afines a los colombianos: una revisión de casos. Sotavento mba, vol. 22, pp. 126-136.
2. Azoumana, K. 2013. Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. Pensamiento Americano, vol. 6, pp. 41-51.
3. Ballesteros, A., Sanchez-Guzman, D., Garcia, R., 2013. Minería de datos educativa: Herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. Lat. Am. J. Phys. Educ., Vol.7, No. 4.
4. Caraveo, L., Sepúlveda, C., 2009. Aplicación de minería de datos para la detección de anomalías: un caso de estudio. Workshop internacional EIG2009.
5. Castoreña, J.A., Silva, A.E, Domínguez, A.J., Rodríguez, D.L., 2018. El uso de herramientas tecnológicas de minería de datos en el análisis de datos climatológicos.

- Revista iberoamericana de las ciencias computacionales e informática, Vol. 7, num. 13.
6. Herrera Santi, P., 1999. Principales factores de riesgo psicológicos y sociales en el adolescente, *Revista Cubana de Pediatría*, 71(1), pp. 39–42.
  7. Malberti M., Klenzi R., Beguerí G., 2017. Análisis, interpretación y toma de decisiones estratégicas en la Ciencia de los Datos. Workshop de investigadores en Ciencias de la computación RedUNCI2016.
  8. Marz, N., Warren J., 2015. *Big Data: Principles and best practices of scalable realtime data systems*. Primera Edición. Greenwich, CT USA.
  9. Mitra, S., Acharya, T., 2003. *Data Mining: Concepts and Algorithms From Multimedia to Bioinformatics*, John Wiley & Sons, Inc.
  10. Peña, A. A., 2014. Educational Data Mining: A Survey and a Data Mining-Based Analysis of recent Works. *Expert Systems with Applications*, vol. 41, pp. 1432- 1462.
  11. Pita, G. E. C. (2018). Las TICs en las empresas: evolución de la tecnología y cambio estructural en las organizaciones. *Dominio de las Ciencias*, 4(1), 499-510.
  12. Reyes-Nava, A., Allan F., Alejo R., Rendón-Lara, E., 2017. Minería de datos aplicada para la identificación de factores de riesgo en alumnos. *Research in Computing Science*, vol. 139, pp. 177-189.
  13. Rico, A., Sanchez, D., 2018. Diseño de un modelo para automatizar la predicción de rendimiento académico en estudiantes del IPN. *Revista iberoamericana para la investigación y el desarrollo educativo*, vol. 8, no. 16.
  14. Ruiz, J., 2011. Herramientas para la investigación en tecnologías de la información y la comunicación. Casos de estudio. *Revista de Currículum y Formación del Profesorado*, 15(1), pp. 139-149.
  15. Sharma, A., Sharma, R., Sharma, V., Shrivatava, V., 2014. Application of data mining – a survey paper. *International Journal of Computer Science and Information Technologies*, vol. 5(2), pp. 2023-2025