

# MODELO DE GESTIÓN TIC APOYADO EN LA MINERÍA DE DATOS COMO ESTRATEGIA DE PREDICCIÓN DE LA DESERCIÓN

Francelina Romero Reyes<sup>1</sup>

Erika María Sandoval Valero

*Universidad Nacional Abierta y a Distancia —UNAD—*

## Resumen

En este artículo se propone el modelo basado en árboles de decisión con algoritmo CART para predecir la deserción escolar, como proyecto de investigación en la Maestría Gestión de Tecnología y de la Información. Este modelo de gestión se basa en la implementación del análisis que involucra actores representativos en el proceso escolar (estudiantes, docentes, coordinadores, psicorientadora, rector y secretaria), con el objetivo de encontrar factores influyentes en la deserción escolar de los estudiantes en el nivel de secundaria, utilizando como variable relacionada con las estrategias de permanencia, prácticas pedagógicas, la motivación académica y el sistema de evaluación. Se utilizará un sistema híbrido basado en herramientas de análisis de infraestructura de negocios, empresarial y de solución TI, apoyadas en técnicas para minería de datos desde una visión metodológica que determine el nivel de deserción en los últimos 3 años (2018, 2019 y 2020).

**Palabras clave:** minería de datos, estrategias, deserción, predecir, permanencia, algoritmo, árboles de decisión.

## 1. Introducción

El sistema educativo colombiano ha mostrado preocupación en los últimos 5 años por el alto porcentaje de deserción escolar en las instituciones de nivel primario, secundario y media. Se estima una tasa de deserción anual, para el año 2018 de 2,07% y para el año 2019 de 3,08%, según el informe emitido por el semáforo escuela Mineducación (2019). Cada año el Ministerio muestra un reporte con variables por entidades territoriales e instituciones ubicadas en zona urbana y zona rural, determinando probabilidades de deserción a final de cada año sin posibilidad de evitar esta problemática. El modelo de gestión basado en árboles de decisión con algoritmo CART, determina la probabilidad de que un estudiante presente la posibilidad de desertar en un determinado tiempo de escolaridad. La aplicación de técnicas de minería de datos permite, entre otras cosas, predecir cualquier fenómeno, con un porcentaje alto de confiabilidad (Timar & Jim, 2015). En este sentido, la estrategia de la minería de datos estima para la aplicación la técnica de clasificación mediante la metodología inductiva descrita en cinco etapas: selección, procesamiento, análisis y transformación, minería de datos y evaluación para luego presentar resultados que viabilicen la aplicación del modelo de gestión TI y su ejecución. El documento se estructura generando un modelo predictivo en algoritmo CART de clasificación y regresión encontrando patrones de comportamientos, fijando variables independientes factibles en los factores que inciden en la deserción escolar en los dos niveles primaria y secundaria de la Institución Educativa de Coyongal.

## 2. Modelo de gestión TI basado en árboles de decisión

La Institución Educativa de Coyongal es de carácter oficial, ubicada en la zona rural del municipio de Magangué, a orillas del río Magdalena, considerando éste como factor que incide en el proceso académico de los estudiantes debido a la distancia y dificultad de transporte, no cuenta con herramientas didácticas y/o

---

<sup>1</sup> Vicerrectoría Académica y de Investigación. fromerore.77@gmail.com

oportunidad de asistencia a otros programas que inviten al estudiante a valorar la permanencia en su proceso académico y formativo. Uno de los problemas es el factor social, económico y familiar. Por lo que, la distancia de sus viviendas para ir a la escuela es lejana, el servicio de transporte escolar no es garantizado, el desorden alimenticio de los estudiantes por cumplir con un horario establecido para tomar el transporte escolar, las constantes lluvias debido al cambio de clima y la debilidad existente en el sistema de evaluación. Estos son factores que se concluyen después de un breve análisis mediante la técnica metodológica de la observación directa y problemática común del entorno.

La evaluación de la deserción escolar, se realiza a través de la minería de datos documentales como es el resumen del Índice Sintético de la Calidad Educativa (ISCE) y sus respectivos componentes desde el 2015, como también la Meta de Mejoramiento Anual (MMA) a alcanzar en el 2018, muestran los resultados a nivel nacional, territorial e institucional. Teniendo en cuenta los resultados del análisis con cada uno de los componentes utilizados como variables, se propone el modelo de gestión TI apoyado en la minería de datos, basado en árboles de decisión con algoritmo CART. En primera fase, se comienza la construcción de la predicción con la variable *rendimiento académico*, con una población estudiantil matriculada en el año 2018 de 1.083 en los niveles de primaria y secundaria. Se hace la extracción y preparación de la información a través de la estructuración de datos. En segunda fase se constituye la aplicación de la minería de datos por medio de la formulación e interpretación de base de datos del SIMAT 2018 y 2019. En tercera fase, se procesa el modelamiento de árbol de decisión con la herramienta Weka para identificar las variables mediante la visualización de algoritmos para análisis de datos y modelado predictivo unidos a la interfaz del SIMAT y plataforma PuntoEdu.

## **2.1 Árbol de decisión**

Un árbol de decisión es un diagrama que contiene: un nodo raíz donde se encuentran todas las observaciones; nodos internos que albergan a los nodos de división y los nodos hoja que contiene la clasificación final para un conjunto de observaciones (Khalilian *et al.*, 2011). Los árboles de decisión son parte de las técnicas de minería de datos (Márquez-Vera, Cano, Romero, & Ventura, 2012). Un árbol representa una segmentación de los datos, que se crea mediante la aplicación, de una serie de reglas simples (Márquez-Vera, Cano, Romero, & Ventura, 2012). Cada regla asigna una observación, a un segmento basada en el valor de una entrada. Una regla se aplica después de otra, dando como resultado una jerarquía de segmentos dentro de segmentos. La jerarquía se llama árbol y cada segmento se llama nodo (Márquez-Vera, Cano, Romero, & Ventura, 2012). Así, los nodos internos de un árbol representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones, y los “nodos hoja” representan las clases (Amaya Torrado, Barrientos Avendaño, & Heredia Vizcaíno, 2014).

## **2.2 Algoritmos CART**

El algoritmo CART de Leo Breiman (Alcover, Benlloch, Blesa, Calduch, Celma, Ferri, & Robles, 2007), realiza particiones binarias, con el objetivo de que la media de cada rama sea diferente y, por tanto, discrimine con suficiente precisión, un número adecuado de particiones, para asignar a cada hoja un valor cercano a la media de los elementos que caen en ella. Este algoritmo genera árboles de fácil interpretación con resultados óptimos (Ara *et al.*, 2015), lo que se considera una ventaja, al crear modelos predictivos (Lin, 2015).

## **2.3 Herramienta Weka**

Es una plataforma de software para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato. 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. Es considerada su utilización como minería de datos por su alto nivel de funcionamiento en arquitectura para modelar algoritmos implementados en otros lenguajes de programación, soporta varias tareas estándar de minería de datos, especialmente, procesamiento de datos, clustering, clasificación, regresión, visualización, y selección, también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC (*Java Database Connectivity*) y puede procesar el resultado devuelto por una consulta hecha a la base de datos.

#### **2.4 Plataforma SIMAT**

(Sistema Integrado de Matrícula), es una herramienta que permite organizar y controlar el proceso de matrícula en todas sus etapas, así como tener una fuente de información confiable y disponible para la toma de decisiones. Es un sistema de gestión de la matrícula de los estudiantes de instituciones oficiales que facilita la inscripción de alumnos nuevos, el registro y la actualización de los datos existentes del estudiante, la consulta del alumno por institución y el traslado a otra institución, entre otros. Es de fácil análisis y de aplicación de la minería de datos para determinar el índice de deserción y gestionar el modelado del árbol de decisiones con algoritmos CART. La Institución Educativa de Coyongal, obligatoriamente utiliza esta base de datos, la cual representa el total de la población matriculada en la institución.

#### **2.5 Plataforma Punto Edu**

Es una plataforma de fácil acceso de organización administrativa que controla y ejecuta las actividades de rendimiento académico, la asistencia y nivel de desempeño, determinando el sistema de evaluación de la institución.

#### **2.6 Estado del arte**

El modelo de gestión TIC es un sistema híbrido basado en herramientas de análisis de infraestructura de negocios, apoyadas en técnicas para minería de datos desde una visión metodológica que determina los factores relacionados en la deserción de los estudiantes en el nivel de secundaria. Se trata de aplicar la técnica de minería de datos como estrategia para detectar el nivel de deserción en los últimos cuatro (4) años en la Institución Educativa de Coyongal.

La deserción de los estudiantes en las instituciones educativas oficiales, es un componente que ha preocupado al Gobierno nacional, debido a factores que deterioran el proceso de las prácticas docentes por la disminución de estudiantes en el aula. Por tal razón, el Ministerio de Educación Nacional diseñó el Índice Sintético de Calidad Educativa y en éste agregó un componente denominado “deserción”, para analizar la cantidad de desertores por años en cada institución a nivel municipal y nacional. Con este objetivo se plantea en cada año la presentación de las pruebas Saber en el grado noveno y las pruebas del ISCE cada año.

En cuanto a la *deserción* se encontraron referentes que involucran la técnica de la minería de datos como apoyo para la predicción de la deserción, como es el autor Ordoñez (2013), quien plantea un modelo de minería de datos aplicando la metodología CRISM-DM contando con la ayuda del análisis de la información que los estudiantes proporcionan a las bases de datos del sistema académico Syllabus y el entorno de aprendizaje, para lograr obtener patrones de comportamiento para conocer cuáles son las posibles causas que motivan a los estudiantes a desertar de sus estudios de primera etapa universitaria.

Es muy relevante este referente, ya que, utilizan una metodología que permite crear el modelo de gestión TIC para la generación de proyectos de minería de datos, pretendiendo obtener un modelo de análisis de datos con la ayuda de la implementación del algoritmo de inteligencia artificial ya incorporado en la herramienta de procesamiento de datos Weka para conocer las causas de la deserción en la primera etapa universitaria. La autora Escobar (2015) en su monografía utiliza herramientas para identificar los factores asociados a la deserción escolar en la Institución Educativa Superior María Inmaculada en el nivel de secundaria, periodo escolar 2014.

Esta monografía se refiere a la construcción de una primera evidencia de la emergencia de conductas negativas acompañadas por una trayectoria educacional marcada por la indisciplina, la inasistencia frecuente a clase, el desinterés y bajo rendimiento académico, es por eso que se requiere indagar la deserción en esta institución y así ayudar a solucionar en gran parte esta situación por la que atraviesan muchos jóvenes.

Según Carvajal (2012) propone estrategias que reduzcan el impacto de las causas académicas en la deserción. También divide la deserción en tres tipos: la deserción precoz, temporal y tardía señalando al estudiante de tres maneras: el desertor del grado, el desertor de la institución y el desertor del sistema educativo. Basándose en estas categorizaciones, se facilita el análisis y la selección de datos para predecir la deserción en un grado en específico, como es en el caso de noveno del nivel de secundaria.

Otro aporte referencial es el de Varón (2017), este reporte “analiza el problema de la deserción estudiantil en la Institución Educativa Siete de Agosto, de Cali, por medio del estudio de los factores internos y externos y el análisis del desarrollo de los procesos educativos como causantes del retiro escolar”.

En este artículo se sugieren también estrategias encaminadas a disminuir el riesgo de desertar mediante la descripción, realizando un estudio exploratorio y utilizando el análisis cuantitativo y cualitativo de la comunidad educativa relacionada con este fenómeno educativo, con un especial interés por el ser. Con esta estrategia se busca identificar una tendencia de retiro escolar con diversidad de variables en un periodo de tiempo específico dentro del marco temporal seleccionado.

### **3. Metodología**

Se busca alcanzar metas que permitan conocer resultados verídicos mediante la recopilación de fuentes primarias y secundarias en el desarrollo de cada uno de los procesos a través de la implementación de métodos de investigación como el método de análisis y síntesis como también el inductivo – deductivo, los cuales conllevan a la realidad y así obtener resultados que permitan ofrecer alternativas de solución al problema encontrado.

La propuesta es importante porque se hace pertinente diseñar estrategias para realizar un análisis exhaustivo y encontrar el nivel de deserción mediante un modelo de gestión TIC, a través de los resultados del SIMAG, pruebas ICSE y pruebas saber entre los años 2015 a 2018.

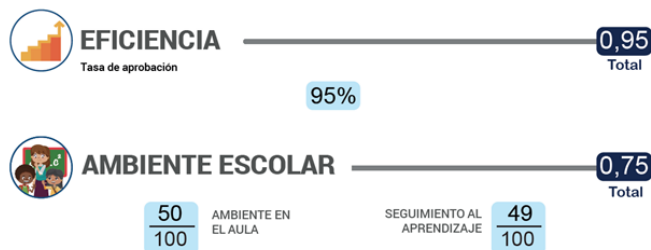
Para el desarrollo de la investigación se revisaron más de 10 fuentes bibliográficas, con el fin de recopilar información y acumular antecedentes correspondientes a la selección y agrupación de datos para predecir la deserción mediante la implementación de gestión TIC, apoyado en la minería de datos como estrategias con los estudiantes en el nivel de secundaria en la INSTEC.

El análisis se realizará involucrando actores representativos en el proceso (estudiantes, docentes, coordinadores, psicorientadora, rector y secretaria) dando origen a la deserción como variable relacionada con las estrategias de permanencia, practicas pedagógicas, la motivación académica y el sistema de evaluación. Se realizará en dos fases:

Primera fase: se hará recolección y selección de la información para organizar los datos pertinentes a la metodología usada, utilizando estrategias como la minería de datos.

La segunda fase: se procede a utilizar las estrategias de infraestructura Pestel para predecir la deserción de los estudiantes del nivel de secundaria.

Continuando el proceso y desarrollo de la primera fase de realiza el análisis del reporte de la excelencia de los años 2017 y 2018 en donde se muestran las metas alcanzadas de las propuestas planteadas para mejorar la calidad educativa en la Institución Educativa de Coyongal. El resumen del Índice Sintético de la Calidad Educativa (ISCE) y sus respectivos componentes de eficiencia y ambiente escolar, desde el 2015, como también la Meta de Mejoramiento Anual (MMA) a alcanzar en el 2018, muestran los resultados a nivel nacional, territorial e institucional, aplica la minería de datos para predecir la deserción escolar.



Cabe resaltar que en el componente eficiencia se muestra una tasa de aprobación del 95%, con un faltante significativo del 5% de reprobación. En el componente ambiente escolar se muestra un seguimiento escolar del 49% dando veracidad al planteamiento que manifiesta debilidad como factor en la deserción escolar en el grado noveno.

#### 4. Conclusiones

El seguimiento del modelo de Gestión TIC apoyado en minería de datos basado en árbol de decisión es una propuesta que ofrece un trabajo administrativo de la Institución Educativa de Coyongal, con el objetivo de predecir la deserción escolar y luego plantear estrategias de permanencias a tiempo con los estudiantes de los niveles de primaria y secundaria.

#### Agradecimientos

Agradezco a mi asesora Erika Sandoval por esta importante orientación en el planteamiento de mi tesis de grado para optar el título de magister en Gestión de la Tecnología y la Información.

#### Referencias

- Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., & Robles, A. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. *XIII Jornadas de Enseñanza Universitaria de la Informática*. <http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/alanal.pdf>
- Amaya Torrado, Y. K., Barrientos Avendaño, E. & Heredia Vizcaíno, D. J. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. <https://documentos.redclara.net/bitstream/10786/759/1/124-22-3-2014-Modelo%20predictivo%20de%20deserci%C3%B3n%20estudiantil%20utilizando%20t%C3%A9cnicas%20de%20miner%C3%ADa%20de%20datos.pdf>
- Márquez Vera, C., Romero Morales, C., & Ventura Soto, S. (2012). Predicción del fracaso escolar mediante técnicas de minería de datos. *Iee-Rita*, 7(3), 109–117. <http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf>
- Pérez López, C., & Santín González, D. (2007). *Minería de datos: técnicas y herramientas*. Madrid: Thomson Ediciones Paraninfo.