

Noviembre
14 -15 -16
2018



VIII CLABES
PANAMÁ - 2018

Octava Conferencia
Latinoamericana
sobre el ABandono
en la Educación Superior

APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA LA IDENTIFICACIÓN DE PATRONES DE DESERCIÓN ESTUDIANTIL COMO APOYO A LAS ESTRATEGIAS DE SARA (SISTEMA DE ACOMPAÑAMIENTO PARA EL RENDIMIENTO ACADÉMICO)

Línea Temática 1: Factores asociados al abandono. Tipos y perfiles de abandono

Calvache Fernández, Leidy Carolina.

Alvarez Vallejo, Valentina

Triviño Arbeláez, Jorge Ivan

Quiceno Restrepo, Claudia

Pulgarin Giraldo, Robinson

Universidad del Quindío – Quindío, Colombia.

jitrivino@uniquindio.edu.co

Resumen. Este artículo aborda el tema de la deserción estudiantil en la educación superior, con la perspectiva de elaborar modelos de Minería de datos que permitan identificar patrones de deserción. Se realizó una revisión sistemática para precisar características que inciden en la deserción universitaria. Para ello se inicia con un estudio de la problemática partiendo desde una revisión literaria de la deserción en diferentes Instituciones de Educación Superior y finalizando concretamente en el Programa de Ingeniería de Sistemas y Computación de la Universidad del Quindío, escenario de estudio. A lo largo de este artículo, se presentan causas de deserción comunes entre los estudiantes, así mismo, se detalla el Proyecto SARA (Sistema de Acompañamiento para el Rendimiento Académico), proyecto constituido a mediados del año 2014 con el propósito de ayudar y acompañar académicamente a los estudiantes en los primeros semestres o aquellos que tiene bajo rendimiento académico, brindando apoyo en temas o en áreas que presentan mayor dificultad de aprendizaje. SARA ha venido consolidando un modelo de estrategias de acompañamiento permanente que incentiva un mejor desarrollo académico y personal, de la mano de Bienestar Institucional, buscando incrementar la motivación y la posibilidad de permanencia de los estudiantes que ingresan al programa. Analizar el abandono universitario da la posibilidad de crear nuevas estrategias, detectar los estudiantes con alto riesgo de deserción es para SARA una labor preventiva muy importante. Observar los síntomas de forma temprana permitirá tomar medidas necesarias y estrategias de prevención en aquellos estudiantes que puedan abandonar el programa académico. El objetivo de este estudio fue incorporar la aplicación de un Proceso de descubrimiento de conocimiento en bases de datos o KDD, para caracterizar y predecir posibles casos de deserción en el programa a través de modelos de minería de datos. Para la construcción de los modelos, se seleccionaron de las bases de datos de la Universidad del

Quindío los datos que relacionan características personales, socioeconómicas y académicas de los estudiantes. Con los datos recopilados para el análisis, se construyó un repositorio que fue procesado y transformado obteniendo así un conjunto de datos depurado y listo para la aplicación de los algoritmos de Minería de Datos. Se descubrieron reglas y perfiles personales, socioeconómicos y académicos de los estudiantes utilizando técnicas de Árboles de decisión.

Descriptor o Palabras Clave: Patrones, Estrategias, Minería de Datos, KDD, SARA.

1. Introducción

Uno de los principales problemas que enfrenta Colombia en cuanto a la educación, concierne a los altos índices de deserción estudiantil en la Educación Superior, según fuentes del Ministerio de Educación Nacional, de cada cien estudiantes que ingresan a la educación superior cerca de la mitad no logra concluir su propósito educativo (Guzmán Ruiz, Durán Muriel , & Franco Gallego, 2009). La deserción se entiende como el abandono escolar por parte de los estudiantes por causa de una combinación de factores como el perfil vocacional que define los intereses, aptitudes, personalidad y capacidades que tiene una persona con respecto a la elección de una carrera universitaria, otros factores como los económicos, familiares, sociales y/o personales hacen parte de este contexto. Se puede definir entonces la deserción estudiantil como el fenómeno que comprende a quienes no siguieron la trayectoria esperada de su programa académico, es decir, la situación a la que se enfrenta un estudiante cuando aspira y no logra concluir sus estudios.

El Ministerio de educación nacional Colombiano realiza un seguimiento a la deserción estudiantil mediante el Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior (SPADIES) que permite medir y monitorear los factores determinantes de la deserción, conocer su evolución en el tiempo y ver cómo se comportan diferentes instituciones y regiones. En el informe de **Estadísticas Deserción y Graduación del 2015** presentado por el Ministerio de Educación Nacional, se registra una tasa de deserción del 9.3% en nivel de formación Universitaria. Por su parte la Universidad del Quindío como lo registra el SPADIES tiene una tasa de deserción del 8.6% al finalizar el año 2015, una tasa superior a la Universidad Nacional de Colombia y la Universidad de Nariño con un índice de deserción del 5.92% y 7.11% respectivamente en el mismo año. El programa de Ingeniería de Sistemas y Computación como un programa adscrito a la Universidad del Quindío no se encuentra al margen de esta gran problemática, al cierre del año 2016 el SPADIES registró para el Programa de Ingeniería de Sistemas y Computación una tasa de deserción del 12.5 %.

2. SARA (Sistema de Acompañamiento para el Rendimiento Académico)

El Sistema de Acompañamiento para el Rendimiento Académico SARA, es un proyecto constituido a mediados del año 2014, diseñado principalmente para lograr la permanencia estudiantil, haciendo uso de estrategias que buscan mejorar el nivel académico, y a la vez, motivar a los estudiantes para lograr una adaptación apropiada y un sentido de pertenencia por sus estudios (Quiceno Restrepo & Pulgarín Giraldo, 2016). El Objetivo principal de este proyecto es ayudar y acompañar académicamente a los estudiantes del Programa de Ingeniería de Sistemas y Computación de la Universidad del Quindío, enfocado principalmente en los

dos primeros semestres de la carrera o aquellos que se encuentren en situación condicional por su bajo rendimiento académico.

Esta población de estudiantes beneficiados por el proyecto, que se menciona anteriormente, fue analizada y soportada con estudios a nivel nacional y particular para el Programa, en donde según el Ministerio de educación “El periodo crítico en el cual el fenómeno se presenta con mayor intensidad, corresponde a los cuatro primeros semestres de la carrera...” (Ministerio de Educación Nacional, 2015) . Particularmente, para el Programa de Ingeniería de Sistemas y Computación de la Universidad del Quindío, como resultado de los análisis realizados para observar el comportamiento de una Cohorte a través de los 10 semestres, se obtuvo que al finalizar el primer semestre desertaron alrededor del 40% de los estudiantes que ingresaron al programa (Garcia Gonzales , Galvis, Hurtado, & Mendez, 2016).

Debido a lo anterior, SARA adopta, construye y maneja estrategias basadas en el acompañamiento continuo a los estudiantes, buscando incrementar la motivación, la pertenencia por la carrera y la posibilidad de permanencia. Estas estrategias, se estudian y se ponen en práctica teniendo en cuenta que su objetivo es, no solo mejorar el nivel educativo de los estudiantes, sino que también se buscan técnicas de estudio adecuadas, para que el estudiante pueda afrontar sus estudios.

La *Fig. 1* permite visualizar algunas de las estrategias manejadas por el proyecto SARA, entre las que encontramos, la Inducción a estudiantes nuevos donde se socializa casos de éxito de estudiantes y egresados para motivar y brindar un sentido de pertenencia, además de brindar una guía sobre los procesos, deberes y derechos que se encuentran enmarcados en el estatuto estudiantil. La Inducción a la biblioteca y Bases de Datos Virtuales y el seguimiento continuo de las pruebas BADyG (Batería de Aptitudes Diferenciales y Generales) son otras de las estrategias implementadas dentro de esta iniciativa.



Fig. 1 Estrategias manejadas por SARA

Entre otras estrategias que SARA ha implementado están, la coordinación para la aplicación de talleres de técnicas de estudio (realizadas mediante el Psico-orientador de Bienestar Institucional) y de charlas sobre aspectos psicosociales y sustancias psicoactivas (realizadas por los psicólogos de Bienestar Institucional). Revisión y análisis de estadísticas de resultados obtenidos de las evaluaciones parciales de primer y segundo semestre. Programación de asesorías grupales o personalizadas de acuerdo con las necesidades de los estudiantes. Remisión temprana de estudiantes que requieren ayuda de Bienestar institucional (psicólogas, psico-orientadores y programas especiales). Además, publicación de información de interés para los estudiantes en su sitio web¹⁰⁵.

3. Metodología

El proceso KDD es una solución al análisis de datos masivos con el objetivo de encontrar información útil que pueda convertirse en conocimiento, es un proceso conformado por una serie de etapas que tiene como principal objetivo la extracción de conocimiento en bases de datos, o conocido también como el proceso no trivial en el cual se pueden identificar patrones válidos, novedosos y útiles a partir de los datos (Hernandez, 2004).

- **Pre-etapa de Identificación de objetivos:** Autores definen esta pre-etapa como el análisis conciso de los procesos donde se identifican los objetivos KDD desde el punto de vista del cliente, adquirir un conocimiento previo del negocio permitirá tener en claro los objetivos que se quieren lograr.
- **Etapas de Selección:** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde las fuentes de datos.
- **Etapas de Preprocesamiento:** Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos, se ponen en práctica operaciones y técnicas necesarias para eliminar ruido, inconsistencias o redundancias de las fuentes de información.
- **Etapas de Transformación:** Esta etapa consiste en adecuar los datos para que sean ingresados a un algoritmo concreto de Minería. Dentro de esta etapa se puede numerizar, normalizar o discretizar buscando adaptar los datos a las necesidades de los algoritmos.
- **Etapas de Minería de Datos:** El objetivo de esta etapa es producir conocimiento, es decir, encontrar relaciones entre los datos o patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles.
- **Etapas de Interpretación y Evaluación:** Durante esta etapa se interpreta y verifica la calidad de los patrones resultantes a través de técnicas de evaluación. Idealmente, los patrones descubiertos deben de tener tres cualidades: ser precisos, comprensibles e interesantes, es decir, útiles y novedosos.

4. Resultados del Proceso de descubrimiento de conocimiento en bases de datos.

¹⁰⁵ <https://sara.uniquindio.edu.co>

Tomando como metodología el Proceso de descubrimiento de conocimiento en base de datos o KDD, se inicia con la pre-etapa de identificación de objetivos, en esta instancia se establecen las metas a lograr al finalizar el proceso. Como parte de la segunda etapa se seleccionaron de las bases de datos de la Universidad del Quindío, los datos que relacionan características personales, socioeconómicas y académicas de los estudiantes del programa. Con los datos recopilados para el análisis se construyó un repositorio dando paso a la etapa de preprocesamiento, se usaron diferentes operaciones y técnicas para la preparación y limpieza de los datos. Como resultado de la etapa de Transformación se obtienen datos listos y adecuados para la aplicación de algoritmos de Minería de Datos. Se descubrieron patrones los cuales fueron interpretados, evaluados, y finalmente usados para soportar la toma de decisiones en el proyecto SARA.

Pre - Etapa de Identificación de Objetivos

El resultado de esta fase son los objetivos definidos para el proceso de extracción de conocimiento, con estos objetivos se determinará los datos que han de usarse dentro de este proceso, así mismo, se analizará si los datos con los que se dispone permitirán alcanzar o no los objetivos previstos. Es importante aclarar que dentro de este estudio se analizan y determinan factores de deserción tanto en la fase de inscripción como en el transcurso de la vida académica de los estudiantes. La fase de inscripción hace referencia al proceso de admisión que el estudiante realiza para ingresar al programa académico, se recopila información personal y socioeconómica del estudiante. El transcurso de la vida académica se refiere al desarrollo universitario del estudiante, donde se analiza su desempeño académico involucrando resultados de parciales, asistencia a asesorías e información de estudiantes que se encuentran en situación condicional.

A continuación, se detalla cada uno de los objetivos definidos para el proceso de minería:

- **Objetivo 1:** Identificar características comunes en el ámbito personal y socioeconómico que permitan predecir si un estudiante que va a ingresar al programa está en riesgo o no de desertar
- **Objetivo 2:** Determinar patrones de los estudiantes con referencia a resultados de parciales, acuerdos y monitorias, para establecer el comportamiento de dichas variables en el fenómeno de la deserción.

Etapa de Selección

El resultado de esta etapa es la generación de dos repositorios como parte de la integración de las fuentes, un repositorio con información personal y socioeconómica de los estudiantes con un total de 1644 registros y 23 atributos, el cual será identificado como REP01 a lo largo de este documento. Por otra parte, se tiene información académica de los estudiantes, con un total de 143 registros y 14 atributos, el cual será identificado como REP02 a lo largo de este documento.

Etapa de Preprocesamiento

Esta etapa es una de las más importantes del proceso KDD, ya que la calidad de los patrones resultantes depende altamente de la calidad de los datos utilizados. Esta etapa es la responsable de obtener datos limpios y de calidad, para lograr este objetivo se deben realizar algunos procedimientos para detectar valores faltantes, datos outliers o cualquier tipo de inconsistencias en los datos suministrados. Atributos como la edad que contenían valores nulos se completaron con la media correspondiente, así mismo, los datos outliers fueron corregidos con el valor correcto.

Etapa de Transformación

La transformación de los datos es la construcción de atributos, la cual consiste en construir automáticamente nuevos atributos aplicando alguna operación o función a los atributos originales con el objetivo de que estos nuevos atributos hagan más fácil el proceso de minería. Para facilitar la extracción de patrones se discretizaron los valores numéricos a valor nominales, como ejemplo se tiene la edad y las calificaciones del estudiante, otros atributos como el régimen de seguridad, nombre de discapacidad, estado civil entre otros se numerización para favorecer el proceso de minería.

Etapa de Minería de Datos

En esta etapa se implementó un árbol de decisión para cada uno de los objetivos propuestos al iniciar el proceso. Los árboles de decisiones resultantes se presentan en forma de árboles n-arios en donde a partir de un nodo padre o raíz se desprende n cantidad de hijos (Ver Fig. 2). Los árboles de decisión fueron implementados en la herramienta Data Mining Tool Orange, las reglas más relevantes con una probabilidad de ocurrencia mayor al 80% se muestran en la siguiente sección.

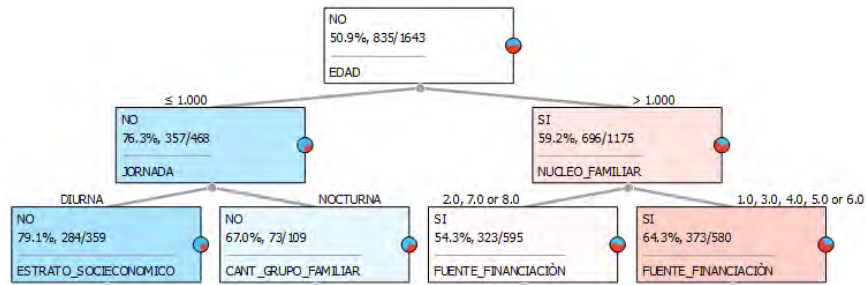


Fig. 2 Árbol de decisión Objetivo 1

Etapa de Interpretación y Evaluación

La evaluación de los modelos se realizó por medio de componentes brindados por Orange, obteniendo una precisión del 70.2% y 80%. Como resultado de interpretar los árboles de decisión generados a partir de los repositorios **REP01** y **REP02**, se obtuvieron las reglas de clasificación más relevantes como se muestra en la *Tabla 1* para el Objetivo 1 y en la *Tabla 2* para el Objetivo 2.

Como se puede observar en la *Tabla 1* los factores de deserción más relevantes en cuanto al ámbito personal y socioeconómico hacen referencia al núcleo familiar, la cantidad de integrantes que lo conforman y la fuente de financiación. Particularmente la probabilidad de

deserción es más alta cuando no se vive con; Dos padres y hermanos, Solo sus dos padres y Sus hijos, o cuando su fuente de financiación involucra Becas de la Universidad, Entidades financieras o Usted y su esposa, adicionalmente que el grupo familiar este compuesto por más de 3 personas.

Tabla 13 Interpretación de reglas Objetivo 1

REGLA	INTERPRETACIÓN	CLASIFICACIÓN	PROBABILIDAD
1	<p>Estudiantes cuya edad es mayor de 20 años</p> <p>Que no vivan con alguno de los siguientes miembros:</p> <ul style="list-style-type: none"> ➤ Dos padres y hermanos, ➤ Solo sus dos padres ➤ Sus hijos <p>Que sus gastos no sean financiados por:</p> <ul style="list-style-type: none"> ➤ Becas de la Universidad ➤ Entidades financieras ➤ Usted y su esposa <p>Que su estrato Socioeconómico sea diferente de 1 y 4</p> <p>Que la cantidad de ingresos sea alguno de los siguientes:</p> <ul style="list-style-type: none"> ➤ Entre 2 y 5 (S.M.M.L.V) ➤ Entre 5 y 10 (S.M.M.L.V) ➤ Mayores a 10 (S.M.M.L.V) <p>Que su categoría del Sisbén sea; 3,4 o 5</p> <p>Que la cantidad de personas que conformen su núcleo sea mayor a 3 personas.</p>	Deserta	100%
2	<p>Estudiantes cuya edad es mayor de 20 años</p> <p>Que no vivan con algunos de los siguientes miembros:</p> <ul style="list-style-type: none"> ➤ Dos padres y hermanos, ➤ Solo sus dos padres ➤ Sus hijos <p>Que sus gastos sean financiados por:</p> <ul style="list-style-type: none"> ➤ Becas de la Universidad, ➤ Entidades financieras ➤ Usted y su esposa <p>Que la cantidad de integrantes de su familia sea mayor e igual a 4 personas.</p>	Deserta	94.1%
3	<p>Estudiantes cuya edad sea mayor a 20 años</p> <p>Que su núcleo familiar este compuesto por alguno de los siguientes miembros:</p> <ul style="list-style-type: none"> ➤ Dos padres y hermanos ➤ Solo sus dos padres ➤ Sus hijos <p>Que su fuente de financiación sea diferente de:</p> <ul style="list-style-type: none"> ➤ Becas de la Universidad ➤ Becas Externas ➤ Icetex ➤ Ingreso de padres y/o familia <p>Que su estado civil sea alguno de los siguientes estados:</p> <ul style="list-style-type: none"> ➤ Separado ➤ Unión Libre ➤ Viudo(a) 	No Deserta	83.3%

En la *Tabla 2* se detallan los aspectos académicos de más relevancia, entre ellos un promedio de notas bajo, particularmente no presentar los parciales del área de Paradigma Orientada a Objetos se ha convertido en el factor de mayor impacto. Adicionalmente cabe resaltar que uno de los factores que garantiza mayor índice de supervivencia es la asistencia a asesorías, un estudiante que asiste al menos a una asesoría de programación garantiza la supervivencia durante su carrera.

Tabla 14 Interpretación de reglas Objetivo 2

REGLA	INTERPRETACIÓN	CLASIFICACIÓN	PROBABILIDAD
1	Aquellos estudiantes que al menos presenten el tercer parcial del área de Paradigma Orientada a Objetos, además una nota mayor o igual a 3,4 en el último parcial de Geometría Analítica	No Deserta	100%
2	Aquellos estudiantes que presenten el tercer parcial del área de Paradigma Orientada a Objetos. Que asista a las monitorias del área de Cálculo en los siguientes temas: <ul style="list-style-type: none"> ➤ Circunferencias y parábolas ➤ Derivadas implícitas ➤ Derivadas. 	No Deserta	100%
3	Aquellos estudiantes que no presenten el tercer parcial del área de Paradigma Orientada a Objetos, además que la nota del segundo parcial del área de Cálculo Diferencial sea menor o igual a 3,4	Deserta	98%

Conclusiones

El Proceso de Descubrimiento de Conocimiento en base de datos permite el manejo eficiente de la información y soportar la toma de decisiones dentro de una organización, brindando la posibilidad de mejorar aspectos o procesos de su entorno. Al detectar información relevante y no trivial, dando un valor diferenciador para la organización. Para el contexto de la deserción estudiantil permite al programa y especialmente al proyecto SARA, conocer cuales variables deberán tenerse en cuenta con mayor prioridad para intervenir en la retención de los estudiantes.

El conocimiento producto de los patrones permitirá generar e incluir estrategias alineadas con la visión y los objetivos que persigue SARA, las cuales podrían estar encaminadas en la intervención de la vida académica de los estudiantes que se perfilan como potenciales desertores, algunas de estas estrategias podrían estar relacionadas con analizar la distribución horaria de asesorías para mejorar ambientes de estudio y crear estrategias motivacionales para que los estudiantes se incorporen con mayor frecuencia en ellas. En cuanto a la aplicación de técnicas de minería de datos en el ámbito educativo, muchos de los patrones encontrados en este trabajo se pueden convertir en un punto de partida para motivar estrategias tempranas de retención estudiantil, relacionadas con la flexibilidad horaria, asesorías académicas, ayudas psicológicas y acompañamiento académico que permitan a los estudiantes continuar con su ciclo educativo.

Referencias

- Garcia Gonzales , M., Galvis, D., Hurtado, L., & Mendez, R. (2016). *Modelo para el seguimiento y control de la deserción en la población universitaria*. Armenia,Quindío: Conceptos Graficos Ltda.
- Guzmán Ruiz, C., Durán Muriel , D., & Franco Gallego, J. (2009). *Deserción estudiantil en la educación superior Colombiana*. Bogotá Colombia.
- Hernandez, J. (2004). *Introducción a la Minería de Datos*. Valencia España.
- Minieducación, & Aprende, C. (2013). *La innovación Educativa en Colombia*.
- Ministerio de Educación Nacional. (2015). *Estrategias para la Permanencia en Educación Superior: Experiencias Significativas*. Bogotá Colombia: Sanmartín Obregón & Cía.
- Quiceno Restrepo, C., & Pulgarín Giraldo, R. (2016). SARA SISTEMA DE ACOMPAÑAMIENTO PARA EL RENDIMIENTO ACADÉMICO. *Encuentro Internacional de Educación en Ingeniería ACOFI*.