



Encuentro Internacional de
Educación en Ingeniería ACOFI

**GESTIÓN, CALIDAD Y DESARROLLO
EN LAS FACULTADES DE INGENIERÍA**

**CARTAGENA, COLOMBIA
18 al 21 de septiembre de 2018**



DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING

Luis Alfredo Mora Osuna, Rodrigo Junior García Hoyos

**Universidad del Sinú
Montería, Colombia**

Resumen

La deserción estudiantil en los programas de pregrado de la gran mayoría de instituciones de educación superior (IES) tanto de Colombia como del resto de Latinoamérica es un problema que tiene un impacto multidimensional en el desarrollo tanto social como económico de un país. Además, en la actualidad, la tasa de deserción de estudiantes en educación superior es uno de los indicadores más utilizados a nivel internacional para evaluar la eficiencia interna de los procesos de enseñanza aprendizaje de las instituciones.

En especial la deserción de primer año, considerando que la mayor fuga de estudiantes se da en ese período. De ahí, se hace necesario estudiarla y buscar unas estrategias que ayuden a solucionar la problemática y que sensibilice a las directivas de la Universidad en la creación de programas de retención estudiantil junto a asesores psicológicos, para que los estudiantes sigan cursando su carrera y no detengan su proceso de formación profesional.

Con esta investigación se quiere desarrollar un sistema que genere perfiles de estudiantes que sean propensos a desertar de la carrera, mediante algoritmos de machine learning, en base al estudio de patrones de datos históricos y recientes del programa de Ingeniería de Sistemas para poder identificar e incluirlos en asesorías psicológicas o en programas de bienestar universitario que ayuden a su acoplamiento estudiantil.

Palabras clave: machine Learning; deserción estudiantil; detección

Abstract

Student dropouts in the undergraduate programs of the clear majority of higher education institutions (IES) in both Colombia and the rest of Latin America is a problem that has a multidimensional impact on the social and economic development of a Country. In addition, at present, the dropout rate of students in higher education is one of the most widely used indicators at the international level to evaluate the internal efficiency of the learning processes of the institutions.

Especialmente la primera defeción, considerando que la mayor escape de estudiantes es dada en ese periodo. Por lo tanto, es necesario estudiar y buscar estrategias para ayudar a resolver el problema y crear conciencia en las directivas de la universidad en la creación de programas de retención de estudiantes con asesores psicológicos, para que los estudiantes continúen persiguiendo su carrera y no dejen su proceso de formación profesional.

With this research we want to develop a system that generates profiles of students who are prone to defect from the race, through machine learning algorithms, based on the study of historical and recent data patterns of the engineering program of systems to be able to identify and include them in psychological counseling or in university welfare programs that help their student linkage.

Keywords: machine learning; student desertion; detection

1. Introducción

La deserción, es el abandono de un programa de estudios antes de obtener el título de pregrado. El número de estudiantes de programas de educación superior casi se duplicó en América Latina durante los últimos 10 años. Sin embargo, dado que apenas la mitad de ellos se gradúan a tiempo, aún queda mucho por hacer en términos de eficiencia y calidad, señaló un nuevo informe del Banco Mundial (Momento decisivo: La educación superior en América Latina y el Caribe, 2017). En Colombia existe un sistema para la prevención de la deserción de la educación superior llamado SPADIES, Este es un sistema para hacer seguimiento sobre las cifras de deserción de estudiantes de la educación superior. Con los datos suministrados por las instituciones de educación superior, se identifican y se ponderan los comportamientos, las causas, variables y riesgos determinantes para desertar.

La deserción es un problema que se viene presentando desde hace varios años en el programa de Ingeniería de Sistemas de la Universidad del Sinú, ha ido incrementándose de forma notoria con una proporción alta de estudiantes que han abandonado sus estudios, por lo general en los primeros semestres de la carrera. Las razones que hay para que un estudiante abandone la universidad son muy diversas, aunque existan causas de deserción que se vuelven muy comunes en la mayoría de los grupos de estudiantes, como causas económicas, familiares o de una mala elección de la profesión, esta situación afecta negativamente a la institución de educación superior ya que ser desertor de la educación

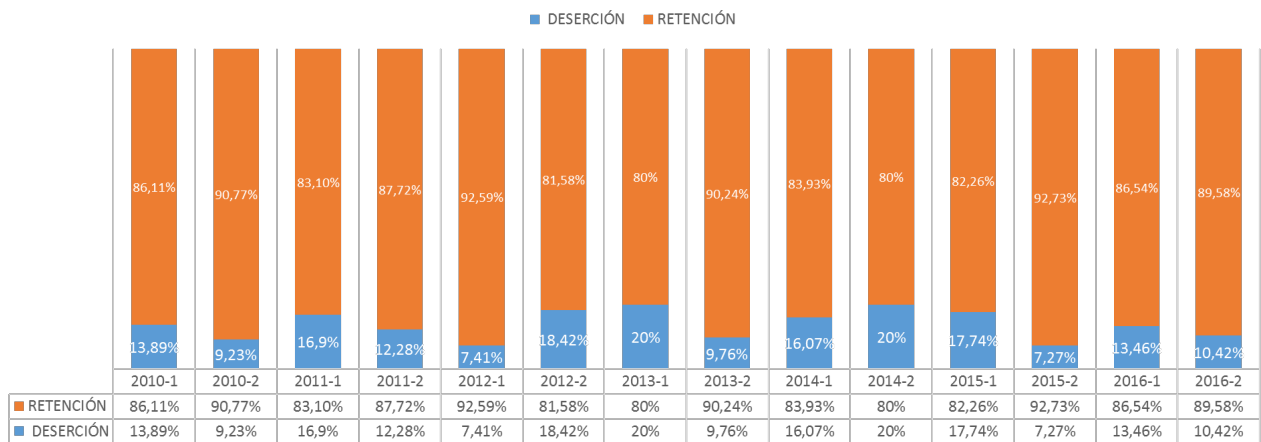
DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING

superior retrasa los avances socioeconómicos y tecnológicos del área donde se presenta este fenómeno.

2. Justificación

La deserción en la educación superior es una problemática analizada constantemente por el Estado mediante el Ministerio de Educación Nacional, y las instituciones universitarias que prestan el servicio de educación. Existe la preocupación porque no siga aumentando este factor en las universidades dado que así lo demuestran las cifras SPADIES para el programa de Ingeniería de Sistemas de la Universidad del Sinú.

DESERCIÓN Y RETENCIÓN POR PERIODO

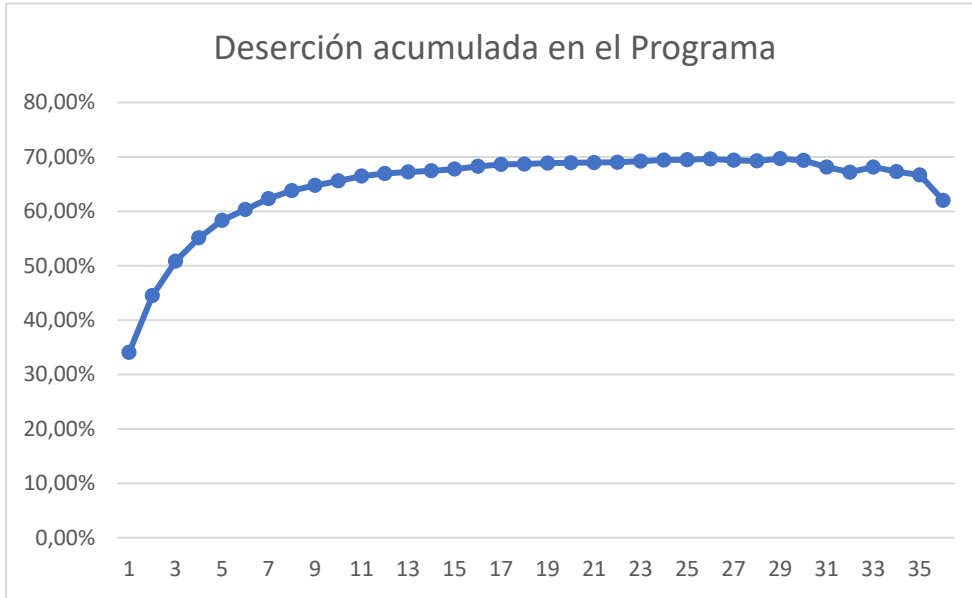


1. Gráfico Tasa de deserción y retención en el Programa (Fuente SPADIES)

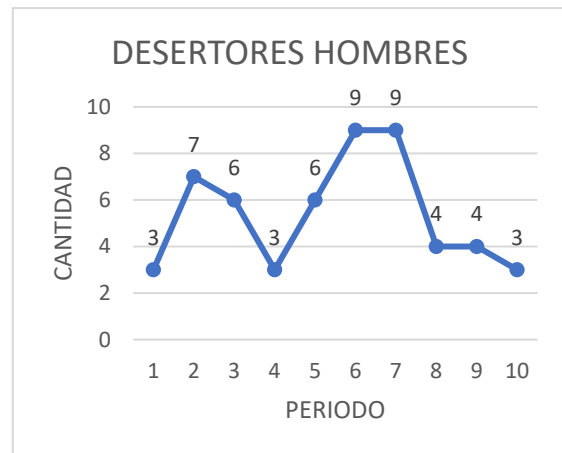
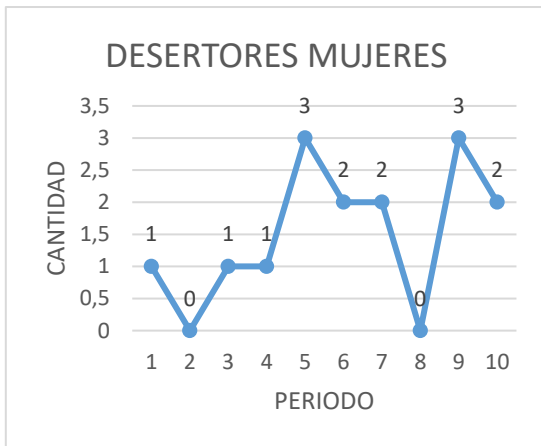
De acuerdo con el gráfico anterior, la deserción en el Programa se ha presentado en un promedio de 13.77% Entre el periodo 2010-1, hasta 2016 – 2; con los puntos más bajos 2012-1 con 7.41% y 7.27% en el 2015-2. Seguidamente la retención se manifiesta en promedio de 86.23%.

La deserción acumulada del programa se muestra en el gráfico 2, donde el mayor índice de deserción se da en el primer semestre del 2013 y el segundo semestre 2014.

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING



2. Grafo Deserción acumulada en el Programa (Fuente SPADIES)
3. Consultados los datos del Programa de Ingeniería de Sistemas de Información registrados en SPADIES, a corte del año 2016-2, se encontró la siguiente caracterización:



En cuanto al género, el podemos observar que, de la población estudiantil del Programa, la mayor deserción está plasmada por los hombres y su pico más alto se dio en los periodos 2014-2 y 2015-1

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING



Por otra parte, la mayor población desertora se dio en el estrato 1, durante el periodo 2015-1; pero mejoró para los periodos subsiguientes, debido a las estrategias del gobierno nacional de apoyar la carrera ingeniería de sistemas, mediante las becas de talento TI.

3. Marco teórico

Machine Learning: disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos

Minería De Datos: el proceso de extraer conocimiento útil y entendible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos.

Almacén De Datos (Datawarehouse): colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

Redes neuronales artificiales: sistema de entradas y salidas compuesto por muchos elementos de procesamiento simples y similares; cada uno de estos elementos tiene un número interno de parámetros los cuales se les denomina "pesos". Al cambiar el peso de un elemento, se altera el comportamiento del mismo y por lo tanto, también cambiará el comportamiento de toda la red.

Regresión logística multiclase: La clasificación mediante regresión logística es un método de aprendizaje supervisado y, por lo tanto, requiere un conjunto de datos etiquetados. Puede entrenar el modelo proporcionando el modelo y el conjunto de datos etiquetados como una entrada a un módulo como Modelo de tren o Hiperparámetros de modelo de sintonización. El modelo entrenado se puede usar para predecir valores para nuevos ejemplos de entrada.

Bosque de decisión multiclase: El algoritmo de bosque de decisión es un método de aprendizaje por conjuntos para la clasificación. El algoritmo funciona construyendo múltiples árboles de decisión y luego votando en la clase de salida más popular. La votación es una forma de agregación, en la que cada árbol en un bosque de decisión de clasificación genera un histograma de frecuencias no normalizadas de etiquetas. El proceso de agregación suma estos histogramas y normaliza el resultado para obtener las "probabilidades" para cada etiqueta. Los árboles que tienen una alta confianza de predicción tienen un mayor peso en la decisión final del conjunto. Los árboles de decisión en general son modelos no paramétricos, lo que significa que admiten datos con distribuciones variadas. En cada árbol, se ejecuta una secuencia de pruebas simples para cada clase, aumentando los niveles de la estructura de un árbol hasta que se alcanza un nodo de hoja (decisión).

K-means: método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano. Es un método utilizado en minería de datos.

4. Metodología

Para la evaluación de los modelos y la verificación de la hipótesis se prepararon una serie de experimentos cuya metodología de validación fue realizada utilizando los parámetros de: exactitud, sensibilidad, área bajo la curva ROC y estadístico de Kappa. Para la evaluación de los algoritmos se ha agregado también la ejecución del algoritmo ZeroR, es conveniente la utilización de este algoritmo debido a su propiedad de clasificar todos los datos en base a la clase mayoritaria, por lo tanto, el porcentaje de aciertos obtenidos por este algoritmo será la meta base a superar.

La metodología de esta investigación va guiada según el desarrollo del proceso de minería de datos que a su vez es regido según la metodología CRISP-DM, la cual está compuesta por las siguientes etapas:

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING

- **ETAPA DE SELECCIÓN:** El objetivo de esta etapa es obtener las fuentes de datos internas que sirven de base para el proceso de elección del modelo de Machine Learning.
- **ETAPA DE PREPROCESAMIENTO DE DATOS:** El objetivo de esta etapa es obtener datos limpios, datos sin valores nulos o anómalos que permitan obtener patrones de calidad.
- **ETAPA DE TRANSFORMACIÓN DE DATOS:** En la etapa de transformación se buscan características útiles para representar los datos según la meta del proceso.
- **ETAPA DE APLICACIÓN DE MODELOS M.L:** El objetivo de esta etapa es la búsqueda y descubrimiento de patrones insospechados y de interés.
- **ETAPA DE INTERPRETACIÓN Y EVALUACIÓN:** En esta etapa se interpretan los patrones descubiertos y posiblemente se retorna a los anteriores pasos o etapas para posteriores iteraciones buscando así el modelo óptimo.

A su vez, se analizó la base de datos histórica con la que cuenta el programa y se lograron determinar una serie de variables que llegan a influenciar positivamente para que se presente el fenómeno de la deserción. Estas variables están divididas en dos grupos:

Grupo 1: Variables económicas, sociales y organizacionales, tomadas de la plataforma institucional ELYSA (PeopleSoft)

Grupo 2: Variables emocionales y psicológicas, tomadas de la plataforma XPLORATE; plataforma que evalúa aspectos tales como: salud física, propósito, bienestar financiero, bienestar emocional y relaciones sociales mediante la aplicación de un test avalado y supervisado por el departamento de psicólogos de la Universidad del Sinú.

Variable	ORIGEN DATOS	Parametrización
Trabaja	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • Si • no
Padres separados	Datos obtenidos de EXPLORATE	No aplica
Relaciones Sociales	Datos obtenidos de EXPLORATE	No aplica
Salud Física	Datos obtenidos de EXPLORATE	No aplica
Bienestar Financiero	Datos obtenidos de EXPLORATE	No aplica
Bienestar Emocional	Datos obtenidos de EXPLORATE	No aplica
Educación de los Padres	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • Algún padre profesional • Ningún padre profesional
Distancia entre la residencia y la institución	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • Cerca • Lejos
Sexo	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • Masculino • Femenino
Promedio acumulado de la carrera	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • 0 - 2,9 • 3.0 – 4.0 • 4.1- 5.0

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTIL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING

Prueba de lenguajes	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • Español • Ingles
Estrato	Datos obtenidos de ELYSA	<ul style="list-style-type: none"> • 1 • 2 • 3 • 4

5. Alcances esperados

- Identificación de variables que pueden influir en la deserción de estudiantes universitarios de primer año.
- Análisis de las causas de la deserción universitaria en el programa de Ingeniería de Sistemas para establecer los mejores predictores de este fenómeno.
- Identificación del algoritmo eficiente para la predicción de la deserción en el programa de ingeniería de sistemas de la Universidad del Sinú.
- Diseño de un software con el algoritmo óptimo en la predicción del nivel de riesgo de deserción que permita generar alertas tempranas de este fenómeno en el programa de Ingeniería de Sistemas de la Universidad del Sinú.

6. Bibliografía

- Ruggeri, F., Kenett, R. S., & Faltin, F. W. (Eds.). (2007). Encyclopedia of statistics in quality and reliability. – Berger, J. B. (2000). Organizational behavior in higher education and student outcomes. HIGHER EDUCATION-NEW YORK-AGATHON PRESS INCORPORATED-, 15, 268-338.
- Angúlo, F., & Sergio, E. (2012). Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios.
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: an overview from a database perspective. IEEE Transactions on Knowledge and data Engineering, 8(6), 866-883.
- Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior. Revista calidad de la educación, 17, 91-108. Torrado, Y. K. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. Red clara. Obtenido de
- Mining, W. I. D. (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- ROJAS, B.M.; GONZÁLEZ, D.C. (2008). Deserción estudiantil en la Universidad de Ibagué. Revista Zona Próxima (No 9), pp. 70-83, ISSN 1657-24, Universidad del Norte, Colombia.
- Timarán, P. R., Calderón, R. A., Ramírez, F. I., Guevara, F., & Alvarado, J. C. (2008). TaryKDD una Herramienta de Minería de Datos Débilmente Acoplada con un SGBD. Memorias de VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento, 3-11.
- Hernández Orallo, J., FERRI RAMIREZ, C. E. S. A. R., & RAMIREZ QUINTANA, M. J. (2004). Introducción a la Minería de Datos. Pearson Prentice Hall,.

DESARROLLO DE UNA HERRAMIENTA COMPUTACIONAL DE ALERTAS TEMPRANAS DE POSIBLES CASOS DE DESERCIÓN ESTUDIANTEL DEL PROGRAMA DE INGENIERÍA DE SISTEMAS DE LA UNIVERSIDAD DEL SINÚ UTILIZANDO ALGORITMOS DE MACHINE LEARNING

Sobre los autores

- **Luis Alfredo Mora Osuna:** Estudiante de ingeniería de sistemas. Luksalfredo19@gmail.com
- **Rodrigo J. García Hoyos:** Ingeniero de Sistemas. Par evaluador reconocido por Colciencias. Docente Universitario. rodrigogarciahoyos@gmail.com

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2018 Asociación Colombiana de Facultades de Ingeniería (ACOFI)