

Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos

Omar D. Castrillón-Gómez, William Sarache y Santiago Ruiz-Herrera

Universidad Nacional de Colombia, Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Industrial, Campus La Nubia Bloque Q piso 2, Manizales - Colombia. (correo-e: odcastrillong@unal.edu.co, wasarachec@unal.edu.co, sruizhe@unal.edu.co)

Recibido Mar. 30, 2020; Aceptado Jun. 2, 2020; Versión final Jul. 10, 2020, Publicado Dic. 2020

Resumen

En este artículo se realiza un estudio sobre las principales variables que afectan el abandono estudiantil. Se predice el comportamiento del abandono estudiantil (variable dependiente) usando 25 variables independientes agrupadas en 5 categorías: personales, económicas, sociales, familiares y académicas. Estas fueron muestreadas sobre una población de 410 estudiantes. Mediante un análisis estadístico multivariado, se seleccionaron aquellas más influyentes para estructurar un archivo que fue analizado por el algoritmo J48 de la plataforma Weka. Los resultados muestran que, con una efectividad superior al 80%, las variables más influyentes en el abandono estudiantil son: pedagogía de los profesores, sentimiento de frustración, importancia del programa, expectativas insatisfechas, motivación del programa y procrastinación. Variables como promedio académico y edad de ingreso parecen no tener mayor relevancia. Se concluye que los resultados aportan información útil para el despliegue de estrategias universitarias orientadas a reducir el abandono estudiantil.

Palabras clave: abandono estudiantil; minería de datos; bayesiano; procrastinación

Prediction of main variables that lead to student dropout by using data mining techniques

Abstract

This research study aims to identify the main variables affecting student dropout. The behavior of student dropout (dependent variable) is predicted. There are 25 independent variables included that are grouped into five categories: personal, economic, social, family, and academic. These variables are sampled from a population of 410 students. The most influential variables are selected by using a multivariate statistical analysis. This generated a file structure that is analyzed using the Weka platform's J48 algorithm. The results show that the most influential variables (effectiveness > 80%) for student dropout are: teacher pedagogy, frustration, the program's importance, unmet expectations, program motivation, and procrastination. Variables such as academic average and age of admission appear not to be relevant. It is concluded that the obtained results provide valuable information for the deployment of university strategies that aim to reduce student dropout.

Keywords: student dropout; data mining; bayesian; procrastination

INTRODUCCIÓN

Hoy en día, no existe un consenso sobre cuáles son las variables más influyentes en el abandono estudiantil. Masserini y Bini (2020) definen esta situación como uno de los problemas más serios generados en la carrera de un estudiante universitario, encontrando que factores como los grupos sociales y redes (por ejemplo Facebook), reducen la tasa de abandono. Hincapié et al. (2012) definen el abandono como el acto de retiro escolar generado por múltiples dimensiones (familiares, sociales, institucionales y personales) y diversas causas como: dificultades de estudio 31%, problemas económicos 22.5%, trabajo 5.8%, servicio militar 3.5% y otros 25%. Lo anterior, lo convierte en un problema muy serio para estudiantes y universidades, el cual (en Colombia) puede alcanzar cifras hasta del 52%. El problema del abandono estudiantil está generalizado en todos los niveles educativos, con tasas superiores al 40% y aunque existe relación con el rendimiento académico, también influye la dinámica familiar (Núñez et al., 2012; Sinchi y Gómez, 2018). Esta problemática puede ser más o menos crítica en dependencia de los sistemas educativos y los contextos sociales y académicos de cada país.

En estudios recientes sobre abandono escolar, llevados a cabo por Vaughn et al. (2020) y Bonilla (2020), se identifican las mayores incidencias económicas, públicas, penales, laborales y comportamentales producidas por este fenómeno, aun cuando el mismo sea temporal. En dichos estudios se destaca la necesidad de determinar los factores que llevan a definir el abandono como un proceso acumulativo de desconexión escolar, recomendándose intervenciones sistemáticas en pequeños grupos desde la escuela, dado que el abandono (temporal o definitivo) en esta etapa, es consistente con el que se presenta en otros niveles de formación. Para estos autores, es igualmente importante motivar a los estudiantes con el fin de cambiar su rol y mejorar sus capacidades. Por tanto, uno de los principales problemas a resolver para la generación de planes de prevención y mitigación del abandono escolar, se centra en la correcta identificación de las causas que lo generan. Establecer dichas causas, es una tarea de vital importancia para identificar aquellos estudiantes con mayor probabilidad de abandonar sus estudios y centrar en ellos acciones de prevención más asertivas.

En el contexto de la educación superior, el problema de abandono ha sido ampliamente analizado. Desde un enfoque social, con el fin de facilitar los procesos de acceso y disminución del abandono, algunas entidades educativas han propuesto alternativas de financiamiento. Sinchi y Gómez (2018) realizaron un estudio bajo la técnica de cuestionario, entrevista y revisión documental a 383 estudiantes, encontrando entre sus principales causas los altos costos y la pérdida de asignaturas. Por su parte, Garzón y Gil (2018) estudiaron la procrastinación académica desde la autorregulación del aprendizaje, estableciendo que este factor influye notoriamente en el abandono y el rendimiento académico. En el mismo sentido, el estudio realizado por Rojas y Gonzales (2008) demostró que la crisis financiera, la falta de orientación vocacional y la edad de ingreso no son los únicos factores influyentes en el abandono estudiantil, pues otros como la amplia oferta educativa también influyen notoriamente, lo cual es más recurrente en los niveles educativos inferiores.

El anterior planteamiento va en contravía con lo establecido por Obergruber y Zierow (2020), quienes en un estudio basado en el análisis de las reformas educativas realizadas en Alemania Occidental entre 1965 y 1996, demostraron que introducir algunos títulos intermedios puede reducir este problema. Otros estudios elaborados en el campo laboral y de salud, realizados a los estudiantes de enfermería, encontraron que los problemas de salud física pueden ser causa del abandono estudiantil (Kox et al. 2020). En el contexto de la educación virtual, Madriz (2016) realizó un análisis de los factores que promueven el abandono en la asignatura métodos I, señalando que lo más importante es evitar el sentimiento de frustración del estudiante. Otros estudios referentes al abandono, en áreas como Massive Open Online Courses (Moocs), muestran el impacto positivo de las estrategias de aprendizaje, basadas en eventos y las variables auto informadas, en la disminución de este fenómeno (Moreno-Marcos et al, 2020).

El problema de abandono también ha sido analizado desde un enfoque matemático-computacional. En el trabajo de Madrid (2017), mediante técnicas estadísticas multivariadas (máquinas de soporte vectorial, análisis Discriminante, K-vecinos más próximos y regresión logística), se clasificaron los estudiantes en aquellos con posibilidad de abandonar y sin posibilidad de abandonar, obteniéndose aciertos promedios del 87%. Este autor, comparó los estudiantes que ingresaron en el año 2009 con los que terminaron en el año 2016, lo cual le permitió obtener un conjunto de datos de entrenamiento y validación. Aunque en dicho trabajo no se identificaron las causas de la deserción, se estableció que el nivel académico al momento de ingreso era una variable influyente en el problema de abandono.

Mediante diferentes modelos de máquinas de aprendizaje entrenados con características académicas, Prieto (2015) encontró, con un nivel de acierto del 85%, los patrones en los factores académicos que aumentan la probabilidad de abandonar un programa. En dicho trabajo se identificó que las suspensiones académicas, las sanciones y el rendimiento académico, eran los factores más influyentes en el abandono estudiantil. Por su parte, Vilorio et al. (2019) realizaron una clasificación bayesiana mediante la herramienta Weka, con el fin de

analizar la probabilidad de abandono de un estudiante, por medio de técnicas de minería de datos. Aunque se lograron tasas de acierto entre el 87% y 91%, no se analizaron las principales variables influyentes en el problema. Usando un red bayesiana, en el trabajo de Denle (2020), con un porcentaje de aciertos del 83%, se identificaron como variables más influyentes en el problema de abandono estudiantil, las siguientes: nivel de interacción social, antecedentes financieros, esperanzas y perspectiva del estudiante. De forma similar, Amaya et al. (2014) desarrollaron un modelo predictivo con base en un árbol de decisión construido por medio del algoritmo J48 del programa Weka. Al comparar los resultados de este programa con la realidad se predice el abandono en un 97%; sin embargo, las variables influyentes no fueron constantes. Así mismo, Timaran (2010), por medio de técnicas de minería de datos, identificó que las variables más influyentes en el abandono se relacionaban con el estrato socio económico bajo, la facultad de ciencias naturales o matemáticas y el rendimiento académico pobre.

Ramírez y Grandón (2018), emplearon árboles de decisión para predecir el abandono académico, obteniendo una precisión del 87.27%. Tales autores identificaron que el promedio de calificaciones, los años de avance en la carrera y el puntaje en la prueba de selección, son las tres variables más influyentes en el problema. Por su parte, mediante técnicas de minería de datos, Timarán y Jiménez (2014) estudiaron los patrones de abandono estudiantil en estudiantes de pregrado, encontrando que las variables más influyentes se relacionaban con el promedio de calificaciones, las asignaturas perdidas, vive con la familia, semestres perdidos, zona de procedencia, ingresos familiares, estado civil, género y tipo de colegio. Igualmente, Pérez et al. (2018) desarrollaron un modelo predictivo de retención de estudiantes durante el primer año. Este modelo identificó relación con variables influyentes tales como la preparación académica, el estrato socioeconómico, los rasgos psicológicos y las características familiares. De esta manera se logró predecir el problema del abandono con un nivel de acierto del 86.4%, usando técnicas de análisis bivariado.

Por su parte, Cruz et al. (2010), al analizar el problema de abandono académico, encontraron que este se relacionaba con el rendimiento académico, los factores económicos, embarazo, depresión, desintegración familiar, estrés, desconfianza, falta de comunicación, adicción, violencia doméstica, falta de respeto, falta de comunicación, influencia familiar al momento de elegir la carrera y cruce de horarios académicos con horarios de trabajo. Dichas variables afectan principalmente la concentración de los estudiantes, así como la realización de sus trabajos y tareas, aspecto que indudablemente influirá en su desempeño académico. Lo anterior, según el citado estudio, produce en los estudiantes tristeza, desmotivación, frustración y, sobre todo, un deseo de abandonar la universidad. Otros estudios relacionados, que proponen estrategias para facilitar la educación y/o disminuir el abandono académico, pueden ser encontrados en Larraín y Zurita (2008), Zacarías y Luna (2018), Palomar-Level y Victorio-Estrada (2017), Rodríguez y Montes (2011), Castañeda et al. (2018), Salazar et al. (2017), Coussement et al. (2020) y Olaya et al. (2020).

Por otro lado, y con miras a mejorar la permanencia escolar, en la actualidad existe una gran variedad de técnicas basadas en herramientas inteligentes, que permiten mejorar los procesos de aprendizaje; entre estas se encuentran (Bharara et al., 2018): el aprendizaje analítico, la inteligencia de negocios, la acción analítica, el análisis de web, la minería de datos educacional, el análisis académico, los sistemas de gestión de contenidos, los recursos educativos abiertos, los sistemas inteligentes de enseñanza, los sistemas de imitaciones, los sistemas de juegos y las técnicas de agrupamiento (K-means, C-means), la lógica difusa, las partículas inteligentes y las cadenas de Markov, entre otras. Estas herramientas permiten descubrir patrones para orientar la toma de decisiones y mejorar los procesos de aprendizaje, con miras a disminuir el abandono académico.

No obstante, y a pesar de la existencia de las contribuciones antes mencionadas, la mayor parte se han orientado a identificar si un estudiante tiene o no potencial de abandonar. Es necesario, por tanto, clasificar a los estudiantes en grupos, según el grado potencial de abandono, desde aquellos con alta probabilidad, hasta aquellos cuya probabilidad de abandonar es casi nula o cero. Esta clasificación, permitirá no solo identificar las principales causas de abandono, sino también definir políticas académicas adecuadas según el nivel potencial de abandono, concentrando mayores esfuerzos en aquellos estudiantes más proclives a hacerlo. Con base en lo anterior, surge la siguiente pregunta de investigación: ¿Cómo clasificar a los estudiantes en grupos, con base en su grado potencial de abandono, a partir de las principales variables influyentes?. Para responder la anterior pregunta de investigación, las principales causas de abandono se agruparon en 4 categorías: personales, económicas, sociales y familiares, y académicas. Cada una de estas categorías se compone de la siguiente manera: a) Personales: motivación por el programa, grado de procrastinación, tiempo estudio, sexo, edad; b) Económicas: problemas económicos, costos del programa, estrato socioeconómico; c) sociales y familiares: importancia de la profesión en el medio, situación cultural, clima familiar, edad de ingreso al programa, tiene hijos, vive con los padres, nivel escolar de los padres; d) académicas: situación académica, expectativas insatisfechas del programa, orientación vocacional, sentimiento de frustración con el entorno educativo, suspensiones o sanciones académicas, número de asignaturas perdidas, colegio de procedencia (público o privado), promedio académico, pedagogía, docentes, relaciones con docentes.

Por tanto, el objetivo del presente estudio es clasificar a los estudiantes en grupos, con base en su grado potencial de abandono (muy alto, alto, regular, bajo o ninguno), a partir de las principales variables influyentes. Para tal fin se utilizó un algoritmo de clasificación bayesiano (J48), ejecutado por medio de la plataforma de aprendizaje automático y minería de datos denominado Weka (<https://www.cs.waikato.ac.nz/ml/weka/>). El aprendizaje automático permite, a partir de diversas técnicas (entre ellas análisis de datos, minería de datos, etc.), entrenar un sistema previamente, para que este pueda predecir el comportamiento de una variable dependiente en una fase posterior, usualmente denominada validación. En este caso concreto, se predice el comportamiento de la variable dependiente denominada abandono, con base en el comportamiento de las variables independientes estudiadas en la investigación. Es importante resaltar que entre más entrenamiento se logró en un sistema, es decir entre más datos existan para realizar la fase de entrenamiento, los resultados serán mejores. Por ende, es necesario contar una muestra estadísticamente válida de los datos.

La plataforma (Weka) permite establecer las principales causas y relaciones que pueden influir en el grado de abandono de un estudiante, mediante un algoritmo que puede ser clasificado como una técnica inteligente (J48). El algoritmo fue aplicado a una base de datos obtenida de una muestra de estudiantes universitarios seleccionados, aleatoriamente y matriculados en el pregrado de Ingeniería Industrial en una universidad pública Colombiana, ubicada en la región central del país. El número de estudiantes seleccionados fue determinado usando métodos estadísticos. Con una efectividad superior al 80%, se encontró que las variables más influyentes en el grado de abandono universitario son: la pedagogía de los profesores, el sentimiento de frustración, la importancia del programa, las expectativas insatisfechas, la motivación que genera el programa y el grado de procrastinación. Estos resultados aportan información valiosa con el fin de que la institución adopte medidas más efectivas contra el abandono estudiantil.

Para su presentación, este artículo ha sido estructurado de la siguiente manera: inicialmente, se expone la metodología empleada para abordar el problema objeto de estudio. Seguidamente, en la sección de resultados, se establecen las principales causas que explican el abandono estudiantil según las salidas arrojadas por el algoritmo de clasificación empleado. En la discusión de resultados se realiza una comparación del estudio frente a otros trabajos similares encontrados en la literatura. Finalmente, las conclusiones exponen los resultados derivados del trabajo y se presenta una lista de las referencias utilizadas.

MATERIALES Y MÉTODOS

Para el diseño y desarrollo de la metodología, se empleó un algoritmo de clasificación bayesiana, el cual fue ejecutado mediante la plataforma de aprendizaje automático y minería de datos denominada Weka (<https://www.cs.waikato.ac.nz/ml/weka/>). Con esta técnica es factible predecir de forma anticipada (*a priori*) el abandono (variable dependiente) que puede tener un estudiante. A partir de esta predicción se aporta información para que la institución diseñe y personalice planes de ayuda para los estudiantes que se pueden encontrar en mayor riesgo de abandono, al determinar las causas más influyentes. Basándose en estudio previos de Castrillón et al. (2020), la metodología empleada en esta investigación, fue estructurada en ocho pasos, así: 1.) Diseño de la encuesta y recolección de la información. 2.) Definición del tamaño de la muestra. 3.) Construcción de la base de datos. 4.) Correlación de variables. 5.) Análisis de componentes principales. 6.) Procesamiento en la plataforma Weka. 7.) Predicción del abandono estudiantil. 8.) Identificación de las principales causas de abandono.

Paso 1: Diseño de la encuesta y recolección de la información. Tomando como referencia los estudios previos de Castrillón et al. (2020) y las principales causas de abandono y categorías en que fueron agrupadas en la introducción, se diseñó una encuesta para evaluar cada una de ellas, usando una escala de cinco niveles: muy alto (5), alto (4), regular (3), bajo (2), ninguno (1). La estructura de la encuesta se expone en la siguiente Tabla 1:

Tabla 1: Categoría y causas del abandono estudiantil

Categoría	Causa	Muy Alto (5)	Alto (4)	Regular (3)	Bajo (2)	Ninguno (1)
Personales	Motivación programa (MP)					
	Grado procrastinación (GP)					
	Tiempo Estudio (TE)					
Económicas	Problemas económicos (PE)					
	Costos programa (CP)					
	Estrato (E)					
Sociales y Familiares	Importancia de la profesión en el medio (IP)					
	Situación cultural (SC)					
	Clima familiar (CF)					
	Nivel escolar de los padres (NEP)					

Tabla 1: Continuación

Categoría	Causa	Muy Alto (5)	Alto (4)	Regular (3)	Bajo (2)	Ninguno (1)
Académicas	Situación académica (SA)					
	Expectativas insatisfechas (EI)					
	Orientación vocacional (OV)					
	Sentimiento frustración con el entorno educativo (SF)					
	Suspensiones o sanciones (SS)					
	Pedagogía (P)					
	Docentes (D)					
	Relaciones con los docentes (RD)					
Deserción	Cuál es la probabilidad de abandonar el programa (PD)					
Sexo (S) : M ____ F: ____ Otro ____						
Edad actual (EA): ____						
Edad de ingreso al programa (EIP): ____						
Número de hijos (NH): ____						
Vive con los padres (VP) : Si: ____ No: ____						
Colegio de procedencia (CP): Público: ____ Privado: ____						
Promedio académico (PA): ____						

Paso 2: Definición del tamaño de la muestra. Las encuestas definidas en el paso uno, fueron aplicadas a un grupo de estudiantes seleccionados, mediante un muestro aleatorio simple. Todos ellos pertenecen al pregrado de Ingeniería Industrial en una universidad pública Colombiana, ubicada en la región central de este país. El tamaño de la muestra fue definido usando la Ecuación (1) (Valdivieso et al., 2011). De acuerdo con dicho autor, el tamaño muestral definido en la ecuación (1) es suficiente cuando se aplican modelos Bayesianos o modelos que se comportan de esta manera.

$$n = \frac{I^2pqM}{ME^2 + I^2pq} \tag{1}$$

Dónde. n: dimensión de la muestra; I: Intervalo de confianza, obtenido por medio de una distribución normal estándar invertida; p y q: representan las variaciones éxito y fracaso, respectivamente; M: Población de estudiantes; E: Error muestral.

Paso 3. Construcción de la base de datos. Con base en el tamaño de muestra, los estudiantes fueron seleccionados al azar. Las encuestas recolectadas y tabuladas se agruparon en 5 categorías, según la escala definida en el Paso 1. Esto, con el fin de clasificar cada una de las causas y el grado de abandono así: Muy alto (5), alto (4), regular (3), bajo (2), ninguno (1).

Paso 4: Correlación de variables. En este paso se establece la correlación existente entre cada una de las causas y la variable dependiente (probabilidad de abandono “PD”). Este paso se realiza, por un lado, para seleccionar las variables con una correlación significativa frente a la variable dependiente y, por otro, para verificar que no existan variables con una alta correlación frente a la variable dependiente.

Paso 5. Análisis de componentes principales (PCA). Con el fin de establecer la influencia de cada variable independiente sobre la variable dependiente y seleccionar aquellas más influyentes, en este paso se realiza un análisis PCA con la ayuda de la librería XrealStats de Excel.

Paso 6. Procesamiento bajo la plataforma Weka. Con base en las causas seleccionadas en el paso anterior, se diseña el encabezado del archivo Arff para la plataforma Weka (<https://www.cs.waikato.ac.nz/ml/weka/>). Este encabezado contempla cada una de las causas analizadas en la encuesta, que fueron seleccionadas en el paso anterior. Adicionalmente, con base en los resultados obtenidos en el paso 3, se estructura el cuerpo del archivo .Arff para ser analizado desde la plataforma Weka.

Paso 7. Predicción del abandono estudiantil. El archivo diseñado y construido en el paso anterior, se analiza usando el clasificador Bayesiano J48. Según Valencia et al.m (2015), esta técnica permite obtener muy buenos resultados con pocos datos. No obstante, en este caso, el análisis puede realizarse con un tamaño de muestra adecuado desde el punto de vista estadístico.

Paso 8. Identificación de las principales causas de abandono. Con base en el algoritmo J48, se obtiene el respectivo árbol de decisión. Este árbol permite identificar las principales causas influyentes en el abandono estudiantil (<http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html>).

RESULTADOS

Al aplicar la metodología antes descrita, se obtuvieron los siguientes resultados, explicados por cada paso del estudio.

Paso 1: Diseño de la encuesta y recolección de la información. Como resultado, de las variables encontradas en esta investigación y referencias literarias estudiadas, en este paso se estructuró la encuesta que se presenta en la Tabla 1 de la metodología. En esta encuesta se consideran 5 categorías principales: Personales, económicas, sociales y familiares, académicas. Sobre cada una de estas categorías se asociaron una serie de variables independientes, y una variable dependiente abandono.

Paso 2: Definición del tamaño de la muestra. Para aplicar Ecuación (1) se usaron los siguientes parámetros: $M = 410$; $l = 1,2815$, intervalo de confianza obtenido para distribución normal estándar invertida (0.9); $p = 0,5$; $q = 0,5$; $E = 0,05$ (error muestral del 5%, para un intervalo del 90%). El tamaño de muestra obtenido (n) fue de 117 estudiantes.

Paso 3. Construcción de la base de datos. Con la información recolectada en el paso anterior, se elaboró la una base de datos sobre la cual se realizó el análisis de deserción (ver Tabla 2):

Tabla 2: Base de datos para analizar el abandono. (Por espacio solo se ilustra una parte)

M P	G P	T E	P E	C P	E P	I P	S C	C F	N E P	S A	E I	O V	S F	S S	P	D	R D	S	E A	E I P	N H	V P	C P	P A	P D
3	3	3	2	3	3	3	4	2	3	4	3	3	3	4	3	3	3	F	21	17	0	S	PU	4,1	4
4	4	4	3	3	3	3	4	3	4	3	3	2	2	2	2	2	4	F	25	18	0	N	PR	3,9	2
4	3	4	3	3	4	5	5	5	5	4	4	4	2	1	3	3	3	M	23	18	0	N	PU	4,4	2
3	4	3	2	3	4	4	4	3	4	3	4	3	3	3	4	4	4	M	19	18	0	N	PR	3,9	3
.
.
.
.
5	3	3	2	3	3	4	4	4	5	4	2	4	2	2	4	4	4	M	20	19	0	N	PR	3,9	1

Paso 4: Correlación de variables. Para medir el grado de relación de las variables independientes respecto a la dependiente, se utilizó el coeficiente de correlación de Pearson. Esta correlación permitió determinar las variables independientes que tienen aporte significativo sobre la variable dependiente (ver variables sombreadas en la Tabla 3). El valor- p para cada una de las correlaciones, en la Tabla 3, se calculó por medio del complemento XRealStats de Excel.

Tabla 3: Correlación de las variables independientes respecto a la variable dependiente

	EI	MP	SF	GP	E	IP	NEP	P	CF	TE	PA	CP	NH
Corre	0,35	0,26	0,25	0,18	0,16	0,095	0,11	0,095	0,09	0,08	0,08	0,06	0,06
Val-p	0,00016	0,0053	0,007	0,058	0,094	0,31	0,245	0,31	0,23	0,4	0,36	0,487	0,501
	PE	SA	CLP	RD	OV	D	EIPN	SS	VP	EAN	SC	S	
Corre	0,06	0,06	0,06	0,05	0,04	0,04	0,04	0,00	0,00	0,00	0,00	0,00	
Val-p	0,501	0,501	0,501	0,65	0,58	0,58	0,58	1	1	1	1	1	

Paso 5. Análisis de componentes principales (PCA). La base de datos de la Tabla 2, fue ordenada por columnas, según el orden de las variables de la Tabla 3. Posteriormente, con la ayuda de la librería XrealStats de Excel, se realizó un análisis de PCA, con el fin de determinar la influencia de cada una de las variables independientes en la variable dependiente. Los resultados de este cálculo se exponen la Tabla 4. En esta tabla se observa que las 8 primeras variables independientes participan en la generación del 65.21% de la información contenida en la variable dependiente. Esto permite deducir que el clasificador que se construya tendrá una eficiencia por lo menos igual ó superior a este porcentaje.

Tabla 4: Análisis de PCA

No	Componente asociado a la variable	Valor del componente	% de influencia	% Acumulado	No	Componente asociado a la variable	Valor del componente	% de influencia	% Acumulado
1	EI	5,20942095	20,84%	20,84%	14	PE	0,61964539	2,48%	84,17%
2	MP	2,37677660	9,51%	30,34%	15	SA	0,59584806	2,38%	86,55%
3	SF	2,15690452	8,63%	38,97%	16	CLP	0,53731090	2,15%	88,70%
4	GP	1,65615753	6,62%	45,60%	17	RD	0,51050045	2,04%	90,74%
5	E	1,38448904	5,54%	51,13%	18	OV	0,45661186	1,83%	92,57%
6	IP	1,35547206	5,42%	56,56%	19	D	0,37690416	1,51%	94,08%
7	NEP	1,09705160	4,39%	60,95%	20	EIPN	0,35487312	1,42%	95,50%
8	P	1,06720757	4,27%	65,21%	21	SS	0,29978950	1,20%	96,69%
9	CF	0,93263822	3,73%	68,94%	22	VP	0,28330638	1,13%	97,83%
10	TE	0,89436410	3,58%	72,52%	23	EAN	0,21776467	0,87%	98,70%
11	PA	0,79872089	3,19%	75,72%	24	SC	0,20619962	0,82%	99,52%
12	CP	0,77936285	3,12%	78,83%	25	S	0,11908164	0,48%	100,00%
13	NH	0,71359821	2,85%	81,69%					

Paso 6. *Procesamiento bajo la plataforma Weka.* En la construcción del archivo para la plataforma WEKA se tomaron las 8 primeras variables de la Tabla 4, pues ellas están asociadas a los componentes que aportan la mayor cantidad de información (65.21%). Con las variables seleccionadas, se construyó el respectivo archivo WEKA (Tabla 5).

Tabla 5: Encabezado archivo weka. @RELATION relation

Campo 1	Campo 2	Campo 3
@ATTRIBUTE	EI	Numeric
@ATTRIBUTE	MP	Numeric
@ATTRIBUTE	SF	Numeric
@ATTRIBUTE	GP	Numeric
@ATTRIBUTE	E	Numeric
@ATTRIBUTE	IP	Numeric
@ATTRIBUTE	NEP	Numeric
@ATTRIBUTE	P	Numeric
@ATTRIBUTE	PD	{"N","B","R","A","MA"}

Para la construcción del encabezado de WEKA, fue necesario convertir cada uno de los valores tomados por la variable dependiente a una denominación literal. Esta conversión se realizó de la siguiente manera: (N = Ninguno =1, B = Bajo = 2, R = Regular = 3, A = Alto = 4, MA = Muy Alto = 5). En la Tabla 6, se ilustra el complemento de los datos para el archivo Weka.

Tabla 6. Datos del Weka: @DATA. Por espacio solo se ilustra una parte de este archivo

V1 EI	V2 MP	V2 SF	V3 GP	V4 E	V5 IP	V6 NEP	V7 P	V8 PD
3	3	3	3	3	3	3	3	A
3	4	2	4	3	3	4	2	B
4	4	2	3	4	5	5	3	B
.
4	3	4	5	3	5	4	3	R
2	5	1	5	2	4	4	4	N
3	4	3	3	5	4	4	4	R

Paso 7. Predicción del abandono estudiantil. El anterior archivo fue interpretado empleando el algoritmo de clasificación bayesiano J48. Cuando se emplea una validación cruzada (90% - 10%), se obtiene un resultado de clasificación del 80% de instancias correctas. Este resultado se expone en las Tablas 7, 8 y en el árbol de clasificación de la Figura 1.

Tabla 7: Clasificación validación cruzada (90% -10%)

Correctly Classified Instances	92	80%
Incorrectly Classified Instances	23	20%
Kappa statistic	0.7078	
Mean absolute error	0.1063	
Root mean squared error	0.2305	
Relative absolute error	37.4741 %	
Root relative squared error	61.3807 %	
Total Number of Instances	115	

Tabla 8: Matriz de Confusión de la Tabla 7

a	b	c	d	e	<-- classified as
26	3	1	1	0	a = N
0	44	4	0	0	b = B
1	4	18	0	0	c = R
1	3	2	4	0	d = A
1	1	0	1	0	e = MA

Es importante resaltar que, cuando se emplea el mismo conjunto de datos para entrenar y clasificar, el porcentaje de clasificación puede llegar hasta el 90% como se ilustra en las Tablas 9 y 10. Un porcentaje un poco más alto.

Tabla 9: Mismo conjunto de datos para entrenar y clasificar

Correctly Classified Instances	104	90.43 %
Incorrectly Classified Instances	11	9.56 %
Kappa statistic	0.8622	
Mean absolute error	0.05	
Root mean squared error	0.1582	
Relative absolute error	17.6447 %	
Root relative squared error	42.1185 %	
Total Number of Instances	115	

Tabla 10: Matriz de confusión de la Tabla

a	b	c	d	e	<-- classified as
28	2	1	0	0	a = N
1	46	1	0	0	b = B
2	0	21	0	0	c = R
1	0	0	9	0	d = A
1	2	0	0	0	e = MA

No obstante, lo ilustrado en las Tabla 9 y 10, los resultados que se usaron en esta investigación son los mostrados en las tablas 7 y 8. De hecho, al comparar tales resultados con otros algoritmos de clasificación similares, se obtienen las eficiencias que se muestran en Tabla 1. De esta manera, se comprueban las bondades del algoritmo de clasificación Bayesiano usado en el presente trabajo.

Tabla 11: Resultados obtenidos por otros algoritmos

Algoritmo	Acierto	Fallo	Referencia
BFTree	61.73	38.26	https://weka.sourceforge.io/doc.packages/bestFirstTree/weka/classifiers/trees/BFTree.html
DecisionStu	44.34	55.65	https://weka.sourceforge.io/doc.stable-3-8/weka/classifiers/trees/DecisionStump.html
FT	75.65	24.34	https://weka.sourceforge.io/doc.packages/functionalTrees/weka/classifiers/trees/FT.html
LadTree	66.08	33.91	https://weka.sourceforge.io/doc.stable/weka/classifiers/trees/LADTree.html
LMT	53.91	46.08	https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/LMT.html
NBTree	48.69	51.93	https://weka.sourceforge.io/doc.packages/naiveBayesTree/weka/classifiers/trees/NBTree.html
RepTree	60.86	39.13	https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/REPTree.html
Simplecart	71.3	28.60	https://weka.sourceforge.io/doc.packages/simpleCART/weka/classifiers/trees/SimpleCart.html

Paso 8. Identificación de las principales causas de abandono. El árbol de decisión de la Figura 1, muestra varias ramificaciones que pueden generar 5 perfiles de estudiantes en riesgo de abandono, de la siguiente manera: a) un alto sentimiento de frustración (SF), mala pedagogía (P) y poca importancia de la profesión (IP), conducen a una probabilidad de abandono media (R); b) una mala pedagogía (P), con un alto porcentaje de expectativas insatisfechas, conducen a una probabilidad de abandono alta (A); c) una mala pedagogía, expectativas insatisfechas (EI), importancia del programa regular (IP), baja motivación (MP) y alto sentimiento de frustración (SF), conducen a una probabilidad de abandono alta (A); d) una mala pedagogía (P), expectativas insatisfechas (EI), importancia del programa regular (IP), motivación del programa regular (MP) y un alto grado de procrastinación (GP), conducen a una probabilidad de abandono media; e) una mala pedagogía (P), expectativas insatisfechas (EI), poca importancia del programa (IP), baja motivación del programa (MP) y un alto grado de procrastinación (GP), conducen a una probabilidad media de abandono.

DISCUSIÓN

Mediante el proceso de clasificación basado en técnicas inteligentes, se identificaron las variables más influyentes en el abandono estudiantil. El anterior estudio muestra que variables como la pedagogía (P), el sentimiento de frustración (SF), la importancia de la profesión en el medio laboral (IP), las expectativas insatisfechas (EI), la motivación en el programa (MP) y la procrastinación (GP), son altamente influyentes en el abandono estudiantil cuando sus valores son medios o altos. Para disminuir los índices de abandono las universidades deben partir de un programa curricular que satisfaga las expectativas insatisfechas de los estudiantes (EI). No obstante, la estrategia debe contemplar otras variables de relevancia tales como una adecuada pedagogía de los profesores (P) y una alta motivación del programa hacia los estudiantes (MP), que evite un sentimiento de frustración alto (SF), que cumpla las expectativas del estudiante (EI) y evite la procrastinación (GP). El control de las anteriores variables podría entonces reducir el grado de abandono de un estudiante. Contrario a lo que se podría pensar, variables como el rendimiento académico y la edad de ingreso no mostraron una influencia notoria en la permanencia de un estudiante en el programa, al menos en la muestra analizada.

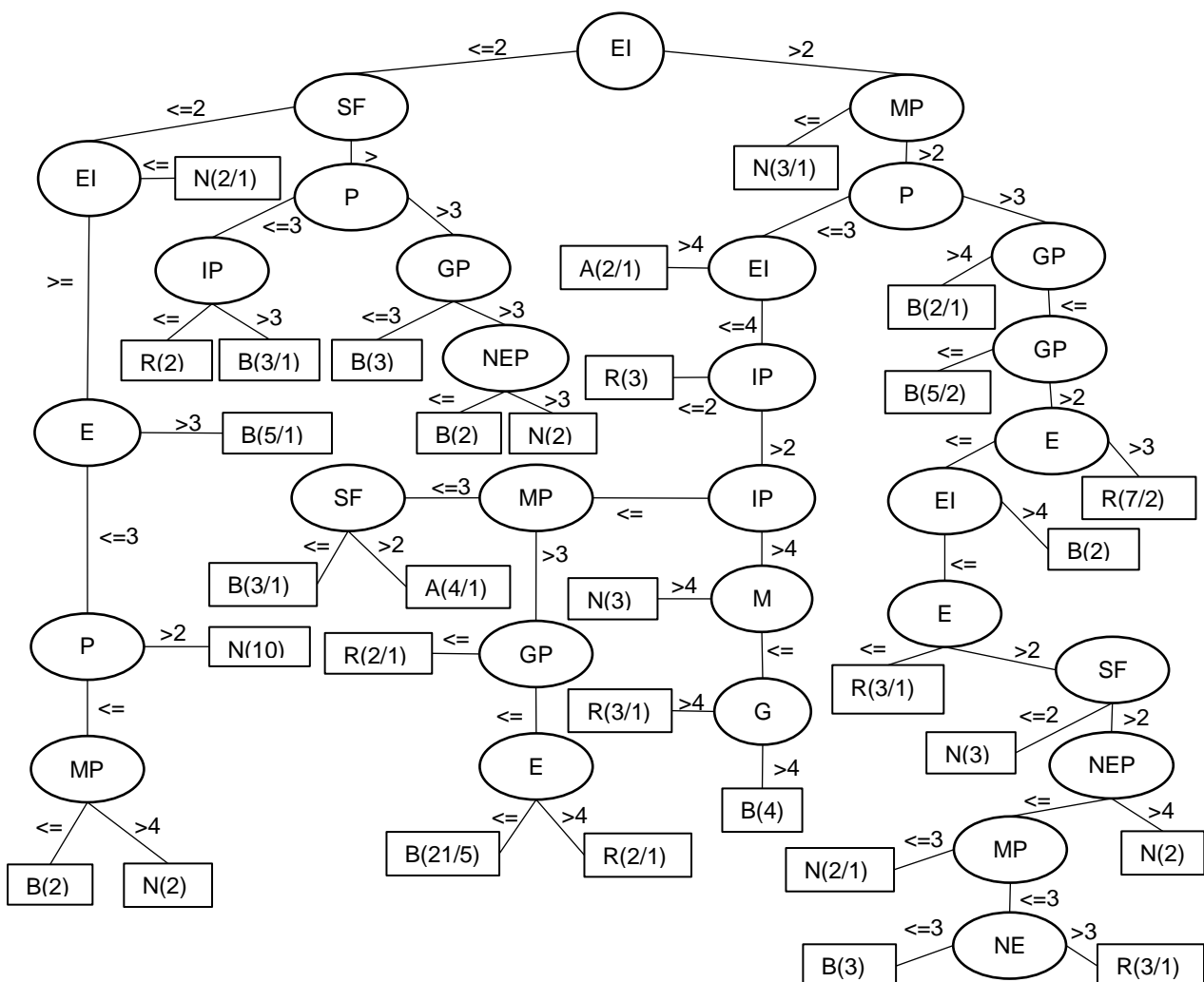


Fig. 1. Árbol de clasificación generado por la plataforma WEKA.

En comparación con la literatura, el presente estudio identifica un conjunto de 6 variables principales e influyentes (pedagogía, sentimiento de frustración, importancia de la profesión, expectativas insatisfechas, motivación en el programa y procrastinación), algunas de ellas no referenciadas en la literatura (p.e. motivación del programa). En contraste, Masserini y Bini (2020) establecen que los grupos y redes sociales son variables tendientes a disminuir el abandono. Igualmente, Hincapié et al. (2012), establece como principales variables de este problema las dificultades de estudio, los problemas económicos, el trabajo y el servicio militar, entre otros. Por su parte Núñez et al. (2012) encuentra como variable predominante la dinámica familiar. Desde otro punto de vista, Sinchi y Gómez (2018) identificaron los altos costos, mientras que Cruz et al. (2010), Ramírez y Grandón (2018), señalan la pérdida de asignaturas como la principal causa de este problema. En concordancia con Garzón y Gil (2018) en esta investigación se encontró que la procrastinación es una variable determinante en el abandono; no obstante, se discrepa en la variable rendimiento académico, respecto a esta última, los resultados obtenidos concuerdan con lo planteado por Núñez et al. (2012), Prieto (2015), Ramírez y Grandón (2018), y Timarán y Jiménez (2014).

Igualmente, la esperanza y las expectativas insatisfechas son variables que coinciden con los hallazgos de Danle (2020). En comparación con Rojas y Gonzales (2008), se discrepa respecto a la variable edad ingreso, pues para tales autores parece ser una variable relevante, mientras que en esta investigación no lo fue. De hecho, Rojas y Gonzales (2008) consideran un conjunto de variables adicionales y diferentes tales como la crisis financiera, la amplia oferta educativa y la orientación vocacional. En el mismo sentido que las investigaciones realizadas por La Madriz (2016), en este documento se identificó que el sentimiento de frustración es una variable influyente en el abandono estudiantil. No obstante, en línea con los planteamientos de Moreno-Marcos et al. (2020), si se incluyen las estrategias de aprendizaje como parte de una adecuada pedagogía, se pueden lograr mejores resultados.

A diferencia del presente trabajo, existen investigaciones que han analizado otras variables relevantes como: obtención de títulos intermedios (Obergruber y Zierow, 2020); problemas de salud física (Kox et al. 2020); nivel académico al momento de ingreso (Madrid, 2017); suspensiones académicas y sanciones (Prieto, 2015); nivel de interacción social, antecedentes financieros (Danle, 2020); estrato socio económico, la facultad de ciencias naturales o matemáticas y mal rendimiento académico (Amaya et al., 2014); años de avance en la carrera, puntaje en la prueba de selección (Ramírez y Grandón, 2018); vive con la familia, semestres perdidos, zona de procedencia, ingresos familiares, estado civil, genero, tipo de colegio (Timarán y Jiménez, 2014); preparación académica, estrato socioeconómico, rasgos psicológicos y características familiares (Pérez et al., 2018); rendimiento académico, embarazo, depresión, desintegración familiar, estrés, desconfianza, falta de comunicación, adicción, violencia doméstica, falta de respeto, falta de comunicación, influencia familiar al momento de elegir la carrera, cruce de horarios académicos con horarios de trabajo (Cruz et al., 2010).

En esta investigación, los diferentes resultados encontrados (80% y 90%) a través de la técnica de minería de datos, se encuentran acordes con los valores comúnmente referenciadas en la literatura. En Madrid (2017), por medio de técnicas de inteligencia artificial, se encontraron aciertos del 87%; sin embargo, los estudiantes fueron clasificados solo en dos grupos, mientras que en el presente trabajo se establecieron 5 grupos, según la probabilidad de abandono. Por su parte, en Prieto (2015), se logró un nivel de acierto del 85%, usando máquinas de aprendizaje. Aunque en la investigación de Viloría et al. (2019), se lograron porcentajes similares a la presente investigación (87% y 91%), usando también técnicas de clasificación bayesiana, no se establecieron las principales variables influyentes. Bajo la plataforma Weka, Amaya et al. (2014), lograron un nivel de acierto del 97%, sin embargo, las variables influyentes no son constantes. Otros trabajos que lograron similares niveles de acierto en comparación con la presente investigación, fueron los de Danle (2020) con 83% usando técnicas bayesianas, Ramírez y Grandón (2018) con 87.27% por medio de árboles de decisión y Pérez et al. (2018) con 86.14% mediante un modelo predictivo (análisis bivariado).

Si bien, esta investigación solo se realiza sobre un programa de pregrado específico, se demuestra que los resultados están acordes a trabajos previos identificados en la literatura, así como las bondades del clasificador Bayesiano como técnica para estudiar el nivel potencial de abandono. Aunque, existes algunas diferencias en los porcentajes de clasificación y las variables influyentes identificadas, esto puede estar ligado a algunos aspectos tales como la población objeto de análisis, los cinco estados de clasificación empleados en este trabajo y el diseño de las técnicas empleadas. De hecho, aquellos autores que han logrado mayores porcentajes de acierto, usualmente no identifican las variables influyentes o estas no son constantes. Es importante resaltar que los resultados son válidos en el caso de estudio analizado, por lo cual es necesario extender la aplicación de la metodología a programas curriculares de otras áreas de conocimiento y a otros contextos geográficos, así como a otros niveles de formación en los cuales podrían surgir otro tipo de causas muy particulares que expliquen el abandono (p.e. programas de doctorado).

Finalmente, se resalta que, dado que esta investigación se enfoca en establecer las causas del abandono estudiantil, no es posible predecir el número de estudiantes que abandonarán un pregrado. No obstante, este aspecto se vuelve de gran importancia en una futura línea de investigación, la cual permita predecir este número y compáralo con los estudiantes que realmente abandonaron el programa.

CONCLUSIONES

Como resultado de este trabajo investigativo se logra: a) identificar las principales variables influyentes en el abandono de un estudiante: pedagogía, sentimiento de frustración, importancia del programa, expectativas insatisfechas, motivación del programa y grado de procrastinación; b) descartar, otras variables como el promedio académico y la edad de ingreso, las cuales mostraron no ser relevantes en este problema; c) una efectividad del 80% y hasta un 90% en el nivel de aciertos de este problema; d) establecer un mínimo número de variables posibles con las cuales se puedan obtener buenos resultados de clasificación, las cuales a su vez, le permitan a las instituciones educativas trazar políticas efectivas para el control de este problema; e) establecer un conjunto completo de 8 variables no analizadas en la literatura (como conjunto), en el análisis de este problema; f) identificar una variable influyente no referenciada en la literatura (motivación del programa) y g) diseñar una metodología efectiva y repetible, para el análisis de este problema.

Finalmente, se resalta que es necesario avanzar en nuevos estudios que permitan observar resultados considerando el nivel de formación (pregrado, maestría, doctorado), la región/país en dónde se ubica la institución, incluyendo otras variables que pudieran ser influyentes en este proceso. Futuras investigaciones deberán orientarse a predecir el número de estudiantes que abandonan un pregrado para realizar comparaciones con el número real de estudiantes que lo terminaron.

AGRADECIMIENTOS

Se agradece la colaboración a la Universidad Nacional de Colombia Sede Manizales y en especial al Departamento de Ingeniería Industrial de esta sede.

REFERENCIAS

- Amaya, K., Avendaño, E. y Heredia, D., *Modelo Predictivo de Deserción Estudiantil Utilizando Técnicas de Minería de Datos*, Universidad Francisco de Paula Santander – Universidad Simón Bolívar, 1-14. Ocaña, Colombia (2014)
- Bharara, S., Sabitha, S. y Bansal, A., *Application of Learning Analytics Using Clustering Data Mining for Students' Disposition Analysis*, DOI 10.1007/s10639-017-9645-7, *Educ Inf Technol*, 3, 957–984 (2018)
- Bonilla, S., *The Dropout Effects of Career Pathways: Evidence from California*, <https://doi.org/10.1016/j.econedurev.2020.101972>, *Economics of Education Review*, 75, 101972 (2020)
- Castañeda, A., Guerra, A. y Ferro, R., *Analysis on the Gamification and Implementation of Leap Motion Controller in the I.E.D. Técnico industrial de Tocancipá*, 10.1108/ITSE-12-2017-0069, *Interactive Technology and Smart Education*, 15(2), 155- 164 (2018)
- Castrillón, O., Sarache, W. y Ruiz, S., *Predicción del Rendimiento Académico por medio de Técnicas de Inteligencia Artificial*, <http://dx.doi.org/10.4067/S0718-50062020000100093>, *Formación Universitaria*, 13(1), 93-102 (2020)
- Coussement, K. y Phana, M., y otros 3 autores, *Predicting Student Dropout in Subscription-Based Online Learning Environments: The Beneficial Impact of the Logit Leaf Model*, <https://doi.org/10.1016/j.dss.2020.113325>, *Decision Support Systems*, 135, 113325 (2020)
- Cruz, E., Barrientos, L. y otros 2 autores, *Academic Performance, School Desertion and Emotional Paradigm in University Students*, *Contemporary Issues in Education Research*, ISSN 1940-5847, 3(7), 25-35 (2010)
- Denle, D., Topuz, K. y Eryarsoy, E., *Development of a Bayesian Belief Network-Based DSS for Predicting and Understanding Freshmen Student Attrition*, <https://doi.org/10.1016/j.ejor.2019.03.037>, *European Journal of Operational Research*, 281, 575–587 (2020)
- Garzón, A. y Gil, J., *El Papel de la Procrastinación Académica como Factor de la Deserción Universitaria*, http://dx.doi.org/10.5209/rev_RCED.2017.v28.n1.49682, *Revista Complutense de Educación*, 28(1), 307-324 (2018)
- Hincapié, F., Hernández, A. y otros 3 autores, *La Deserción Universitaria desde la Perspectiva del Estudiante. Informe de Casos: Programa Microbiología y Bioanálisis*, *Revista Praxis*, ISSN: 1657-4915, 8, 17 – 38, (2012)
- Kox, J., Bakker, E. y otros 4 autores, *Effective Interventions for Physical Health Complaints in Nursing Students and Novice Nurses: A Systematic Review*, <https://doi.org/10.1016/j.nepr.2020.102772>, *Nurse Education in Practice*, article in press (2020)
- La Madriz J., *Factors that Promote the Defection of The Virtual Classroom*, *Revista Orbis*, ISSN: 1856-1594, (35)18 – 40 (2016)

- Larraín, C. y Zurita, S., *The New Student Loan System in Chile's Higher Education*, 10.1007/s10734-007-9083-3, High Educ (55), 683–702 (2008)
- Masserini, L. y Bini, M., *Does Joining Social Media Groups Help to Reduce Students' Dropout Within the First University Year?*, <https://doi.org/10.1016/j.seps.2020.100865>, Socio-Economic Planning Sciences, article in press, (2020)
- Madrid, J., *Propuesta de un Modelo Estadístico para Caracterizar y Predecir la Deserción Estudiantil Universitaria*, Universidad Nacional de Colombia, Medellín, Colombia (2017)
- Moreno-Marcos, P., Muñoz-Merino, P. y otros 4 autores, *Temporal Analysis for Dropout Prediction Using Self-Regulated Learning Strategies in Self-Paced MOOCs*, <https://doi.org/10.1016/j.compedu.2019.103728>, Computers & Education, 145, 103728 (2020)
- Núñez, M., Henao, A. y Fajardo, E., "La Deserción: Un Fenómeno Social". *Vivencias de los Estudiantes del Programa de Enfermería de la Universidad del Tolima*, Revista Praxis, ISSN: 1657-4915, 8, 7 – 16 (2012)
- Obergruber, N. y Zierow, L., *Students' Behavioural Responses to a Fallback Option - Evidence from Introducing Interim Degrees in German Schools*, <https://doi.org/10.1016/j.econedurev.2020.101956>, Economics of Education Review, 75, 101956 (2020)
- Olaya, D., Vasquez, J., y otros tres autores, *Uplift Modeling for Preventing Student Dropout in Higher Education*, <https://doi.org/10.1016/j.dss.2020.113320>, Decision Support Systems, 134, 113320, (2020)
- Palomar-Lever, J. y Victorio-Estrada, A., *Academic Success of Adolescents in Poverty*, 10.1007/s11218-017-9389-7, Soc Psychol Educ, 20, 669–691 (2017)
- Pérez, A., Ruff, C. y otros 3 autores, *Modelo de Predicción de la Deserción Estudiantil de Primer Año en la Universidad Bernardo O'Higgins*, <http://dx.doi.org/10.1590/S1678-4634201844172094>, Educ. Pesqui, 44, 1-23 (2018)
- Prieto, C., *Uso de Regresión Logística para Predecir Deserción Estudiantil Temprana*, Universidad de los Andes, Departamento de Ingeniería de Sistemas y Computación, Bogotá, Colombia (2015)
- Ramírez, P. y Grandón, E., *Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación Basada en Árboles de Decisión con Parámetros Optimizados*, <http://dx.doi.org/10.4067/S0718-50062018000300003>, Formación Universitaria, 11(3), 3-10 (2018)
- Rodríguez, F. y Montes, A., *A Teacher Training Experience Based on Work with Communities in California*, Educación y Educadores, ISSN 0123-1294, 14 (1), 105 -118 (2011)
- Rojas, M. y González, D., *Deserción Estudiantil en la Universidad de Ibagué, Colombia: Una Lectura Histórica en Perspectiva Cuantitativa*, Revista del Instituto de Estudios en Educación Universidad del Norte, ISSN: 2145-9444, 9, 70 - 83 (2008)
- Salazar, M., Lozano, M. y otros 2 autores, *Evaluation of an Academic Satisfaction Model for First-Year University Students*, Journal of Psychological and Educational Research, ISSN: 2247-1537, 25 (2), 115- 140 (2017)
- Sinchi, E., y Gómez, G., *Acceso y Deserción en las Universidades. Alternativas de Financiamiento*, <https://doi.org/10.17163/alt.v13n2.2018.10>, Alteridad, 13(2), 274-287, (2018)
- Timaran, R. y Jiménez, J., *Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM*, Congreso Iberoamericano de Ciencia, Tecnología, Innovación y Educación, ISBN: 978-84-7666-210-6, 1 – 19, Buenos Aires, Argentina (2014)
- Timarán, R., *Una lectura Sobre Deserción Universitaria en Estudiantes de Pregrado desde la Perspectiva de la Minería de Datos*, Revista Científica Guillermo de Ockham, ISSN: 1794-192X, 8 (1), 121-130 (2010)
- Vaughn, M., Greg, R., y otros 3 autores, *Preliminary Validation of the Dropout Risk Inventory for Middle and High School Students*, <https://doi.org/10.1016/j.childyouth.2020.104855>, Children and Youth Services Review, 111, 104855 (2020)
- Valdivieso, C.E., Valdivieso, R. y Valdivieso, O.A., *Determinación del Tamaño Muestral Mediante el Uso de Árboles de Decisión*, DOI: 10.23881/idupbo.011.1-4e, UPB - Investigación & Desarrollo, 11, 148 – 176 (2011)
- Valencia, M., Correa, J. y Díaz, F., *Métodos Estadísticos Clásicos y Bayesianos para el Pronóstico de Demanda. Un Análisis Comparativo*, <https://doi.org/10.15446/rev.fac.cienc.v4n1.49775>, Revista Facultad de Ciencias Universidad Nacional de Colombia, 4(1), 52 -67 (2015)
- Viloria, A., Pineda, O. y Varela, N., *Bayesian Classifier Applied to Higher Education Dropout*, 10.1016/j.procs.2019.11.045, Procedia Computer Science, 160, 573–577 (2019)
- Zacarias, M. y Luna, J., *Teacher Training in Special Education: Study Habits and Teaching Practice*, doi.org/10.17163/alt.v12n2.2017.01, Alteridad Revista de Educación, 13 (2), 255-265 (2018)