



Revista EIA
ISSN 1794-1237
e-ISSN 2463-0950
Año XIX/ Volumen 20/ Edición N.40
Julio - diciembre de 2023
Reia4002 pp. 1-28

Publicación científica semestral
Universidad EIA, Envigado, Colombia

**PARA CITAR ESTE ARTÍCULO /
TO REFERENCE THIS ARTICLE /**

Castro, L.; Espitia, E.; Romero, E.
Análisis de características que
influyen en la deserción estudiantil
en el contexto de una universidad
latinoamericana
Revista EIA, 20(40), Reia4002.
pp. 1-28.
<https://doi.org/10.24050/reia.v20i40.1628>

 *Autor de correspondencia:*


Castro, L.
Universidad del Quindío
Correo electrónico:
luferr@uniquindio.edu.co

Recibido: 04-08-2022

Aceptado: 15-05-2023

Disponible online: 01-06-2023

Análisis de características que influyen en la deserción estudiantil en el contexto de una universidad latinoamericana

 LUIS FERNANDO CASTRO¹
ESPERANZA ESPITIA¹
EDWIN ROMERO¹

1. Universidad del Quindío - Colombia

Resumen

El presente trabajo pretende profundizar en el estudio de la deserción estudiantil universitaria, un problema serio que preocupa a los gobiernos, las instituciones universitarias y estudiantes a nivel mundial. Para lograr lo anterior, este estudio aplica minería de datos con el propósito de analizar la deserción estudiantil en una universidad latinoamericana, basados en el descubrimiento de las características relevantes que tienen mayor incidencia y en la identificación de patrones que faciliten el entendimiento de dicho problema. La metodología empleada se basa en una adaptación de los pasos propuestos por KDD (descubrimiento de conocimiento en bases de datos) y en un diseño de investigación observacional, descriptivo y transversal, con muestreo por conveniencia. La muestra está compuesta por 10705 estudiantes, los cuales se encuentran distribuidos en 7 facultades y 33 programas académicos de pregrado. Las relaciones predictivas entre la condición de estudiante desertor y las características influyentes, se han sometido a verificación mediante un modelo basado en árbol de decisión. Como resultado de este trabajo, se identifican algunas técnicas y métodos comúnmente empleados para este tipo de estudios y se desarrolla un método para identificar patrones de relaciones entre las características más influyentes en el fenómeno de la deserción. Se encontró que las principales características influyentes en este tipo de deserción se refieren a estrato socioeconómico, género, situación laboral y promedio acumulado. Un aspecto a resaltar es la coincidencia de los hallazgos de este trabajo con los resultados de otros trabajos similares a nivel mundial, en los cuales se identificó el rendimiento académico como un factor fundamental que incide en la deserción universitaria. Se concluye que la deserción estudiantil universitaria no depende de una sola característica, sino que es causada por un conjunto de características y su interrelación.

Palabras clave: Investigación de deserción; educación superior; análisis de datos; patrones de relación.

Analysis of characteristics influencing student dropout in the context of a latin american university

Abstract

The present work aims to deepen the study of the student dropout, which is a serious problem that worries the governments, university institutions and students worldwide. To achieve the above, this study uses data mining to analyze student dropout in a Latin American university by discovering the most influential relevant characteristics and by identifying patterns to facilitate the understanding of such problem. The methodology used considers an adaptation of the steps proposed by KDD (knowledge discovery in databases) and the study design was observational, descriptive and cross-sectional, using convenience sampling. The sample is made up of 10705 students, which are distributed in 7 faculties and 33 undergraduate academic programs. A model based on decision tree was used to verify the predictive relationships between the status of dropping out student and the influential characteristics. As a result, this work identified techniques and methods commonly used in these studies and developed a method to identify patterns of relationships among the most influential characteristics in the student dropout. We found that the main influencing characteristics in this study refer to socioeconomic level, gender, employment status and academic performance. One aspect to be highlighted is the coincidence of the findings of this study with the results of other similar studies worldwide, in which academic performance was identified as a fundamental factor that affects university dropout. As a conclusion we can state that university student dropout is caused by a set of characteristics and their interrelations rather than a single characteristic.

Keywords: Dropout research; higher education; data analysis; relationship patterns.

1. Introducción

La deserción estudiantil en instituciones universitarias, definida por Ramírez (2021) como el abandono de un estudiante a un programa de estudio o institución antes de obtener un título o grado académico, es un fenómeno de interés mundial que conlleva importantes repercusiones económicas y sociales no solo para el estudiante sino también para la sociedad en general (Munizaga et al., 2018). En este sentido, Vila (2019) afirma que el abandono estudiantil en la educación superior genera una problemática que impacta no solo aspectos de la economía nacional en términos de desempleo y pobreza, sino también otros aspectos desde el punto de vista social, público y penal. Por otra parte, el Proyecto ALFA-GUIA (2013) describe una revisión general de la temática sobre el abandono estudiantil en la educación superior abordando su estudio, investigación, prevención y atención. Dicho proyecto tuvo

como patrocinador a la Comunidad Europea mediante el programa ALFA, un programa de cooperación entre instituciones de educación superior de la Unión Europea y América Latina. Participaron inicialmente 20 Instituciones de Enseñanza Superior de 16 países diferentes (12 países de América Latina y 4 de Europa) y una entidad colaboradora.

La minería de datos facilita la generación conocimiento a partir de la información almacenada en bases de datos. Según Vila (2019) este conocimiento se obtiene como un modelo basado en la información que se extrae. Zarate (2019), afirma que la minería de datos en el campo de la educación se denomina minería de datos educativa. Es así como, la EDM (Minería de Datos Educativa) emerge como un campo de investigación multidisciplinario para analizar datos educativos aplicando minería de datos en el contexto del descubrimiento de conocimiento en bases de datos comúnmente conocido como KDD (Ayala et al., 2021; Kumar et al., 2017; Zarate, 2019). De acuerdo con Ghazal y Hammad (2022), el descubrimiento de conocimiento en bases de datos (KDD) hace referencia al proceso general de extraer conocimiento nuevo y útil a partir de datos históricos. Para llevar a cabo este proceso, los autores proponen las etapas de entendimiento del problema, selección de predictores, preparación de datos, selección de funciones y algoritmos de minería de datos, evaluación y uso del conocimiento descubierto. Para otros autores, KDD comprende las etapas de recolección de datos, pre-procesamiento de los datos, selección de atributos y aplicación de algoritmos de aprendizaje computacional (UOrbina-Nájera et al., 2020).

Existe una variedad de trabajos relacionados con la aplicación de EDM para el análisis de la deserción estudiantil universitaria los cuales permiten a las instituciones educativas establecer planes de acción conducentes a reducir sus tasas de deserción, mejorar la calidad de la educación y favorecer sus indicadores de desempeño (Bakhshinategh et al., 2018; Castro et al., 2018; Castro et al., 2019; Munizaga et al., 2018). Algunas cifras sustentan la problemática analizada en este tipo de estudios. Según Munizaga et al. (2018), en promedio, la tasa de retención en América Latina y el Caribe es apenas del 46%. Adicionalmente, se tiene que, entre los estados miembros de la Unión Europea, la proporción de abandonos prematuros en 2019 osciló entre el 3.0% en Croacia y el 17.3% en España (Oficina Europea de Estadística [EUROSTAT], 2020). Es así como, uno de los objetivos del Espacio Europeo de la Educación Superior (European Higher Education Area, EHEA) consiste en identificar estrategias para afrontar el fracaso escolar (Clerici & Da Re, 2019; Hatos et al., 2020). Finalmente, de acuerdo con cifras

nacionales en Colombia, para el año 2018, la tasa de deserción anual para los programas universitarios se ubicó en 8.79%, para los tecnológicos en 10.75% y para los técnicos profesionales en 17.41% (Ministerio de Educación Nacional [MEN], 2021).

Fundamentado en lo anterior, en este artículo se desarrolla la hipótesis según la cual el uso de métodos y técnicas basadas en KDD y EDM permiten analizar adecuadamente la deserción estudiantil en una institución universitaria latinoamericana. Los objetivos de este trabajo se orientan a identificar metodologías, técnicas y herramientas que apoyan el proceso de descubrimiento de conocimiento sobre datos, establecer un método que permita analizar la deserción estudiantil basados procesos de descubrimiento de conocimiento sobre datos e identificar patrones de relaciones entre las características de los estudiantes desertores.

2. Materiales y Métodos

Objetivos

El propósito principal de este estudio es analizar la deserción estudiantil en una universidad colombiana, específicamente en la Universidad del Quindío, por medio de procesos de descubrimiento de conocimiento sobre datos. Para lograr lo anterior se identifican algunas propuestas de metodologías, herramientas, algoritmos y técnicas comúnmente empleadas para apoyar el proceso de descubrimiento de conocimiento sobre datos en el ámbito educativo. De esta forma es posible establecer y aplicar un método que permita analizar dicha deserción. Para dar respuesta a los objetivos planteados se optó por un enfoque cuantitativo de investigación basado en Constante et al. (2021) a través de un análisis secundario de datos. Luego se procedió a analizar la información recolectada. El diseño fue de carácter no experimental y se realizó un análisis descriptivo y predictivo, con la finalidad de obtener unos resultados relevantes aplicables a políticas de tipo preventivo y diagnóstico.

Basados en Villalobos (2017), este estudio se clasifica como cuantitativo de tipo exploratorio, no experimental y correlacional, debido a que las características se observan en un período específico, en su contexto natural, sin pretender una manipulación de estas y con interés particular en determinar el grado de asociación entre ellas.

Identificación de metodologías, herramientas, algoritmos y técnicas

Existen diversos estudios que proponen enfoques novedosos para apoyar este tipo de investigaciones. En la Tabla 1 se presenta la relación de algunos estudios relevantes en el área.

Tabla 1 Relación de los estudios revisados		
ID	Referencia	Nombre
ST1	(Oñate, 2016)	Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería de datos.
ST2	(Vásquez, 2016)	Modelo predictivo para estimar la deserción de estudiantes en una Institución de Educación Superior.
ST3	(Villanueva, 2016)	Análisis de deserción en una Institución de Educación Superior Colombiana mediante el uso de Minería de Datos para la Educación (EDM).
ST4	(Ramírez & Grandón, 2018)	Predicción de la deserción académica en una universidad pública chilena a través de la clasificación basada en árboles de decisión con parámetros optimizados.
ST5	(Cuji, et al., 2017)	Modelo predictivo de deserción estudiantil basado en arboles de decisión.
ST6	(Vicente, 2020)	Aplicación de la técnica de minería de datos para la predicción de la deserción estudiantil universitaria.
ST7	(Urbina-Nájera et al., 2020)	Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa.
ST8	(Vila, 2019)	Detección de patrones de deserción estudiantil utilizando técnicas predictivas de clasificación y regresión de minería de datos, para la gestión académica de la Universidad Técnica del Norte.
ST9	(Constante et al., 2021)	Factores asociados al abandono universitario.
ST10	(Castrillón-Gómez et al., 2020)	Predicción de las principales características que conllevan al abandono estudiantil por medio de técnicas de minería de datos.
ST11	(Quiñones et al., 2020)	Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos.
ST12	(Pando & Zarate, 2020)	Aplicación de un modelo de minería de datos para identificación de patrones que influyen en la deserción académica en el instituto superior Leonardo Davinci.

Fuente. Elaboración propia

Dentro de los principales aportes de los estudios anteriores se pueden resaltar las decisiones acerca del conjunto de características que se seleccionaron para los análisis y el descubrimiento de las características que se catalogan como las más influyentes en el fenómeno de la deserción. Dichos aportes se describen en la Tabla 2.

Tabla 2 Relación de características de los estudios revisados

ID	Características analizadas	Características más influyentes en la deserción
ST1	Información socio económica, historial académico y desempeño.	Información socioeconómica e historial académico.
ST2	Sociodemográficas, institución educativa de donde proviene, desempeño académico preuniversitario y universitario, relación con su ambiente universitario y financiamiento.	Desempeño preuniversitario, seguido del nivel educacional de los padres y el desempeño académico universitario
ST3	Socio-vocacional, académico y económico-financiero	Créditos aprobados, promedio programa.
ST4	Demográficas del estudiante, antecedentes académicos, situación socioeconómica, rendimiento académico.	Promedio de notas, años de avance en la carrera y puntaje de ingreso a la universidad.
ST5	Datos personales y académicos, promedio de notas.	Nivel, nota en el primer parcial del segundo período, nota en el segundo parcial del segundo período.
ST6	Carrera, nivel, género, notas obtenidas	Calificaciones pertenecientes al segundo período, calificaciones del primer parcial en tercer semestre, calificaciones del primer parcial en cada semestre.
ST7	Datos demográficos, antecedentes familiares, escolaridad previa, Rendimiento académico actual, apoyos financieros, Ambiente y convivencia, Infraestructura, asesorías, Servicios.	La falta de asesorías, inadecuado ambiente estudiantil, falta de seguimiento académico, deficiente calidad educativa y al servicio en general.

ID	Características analizadas	Características más influyentes en la deserción
ST8	Datos personales, académicos, socioeconómicos, demográficos, psicológicos.	Género del estudiante, el nivel económico del estudiante, el promedio de notas académicas, estado civil, etnia y el lugar de procedencia.
ST9	Demográficas, socio económicas y académicas	Dedicación del estudiante, edad, nota de acceso a la universidad, área de conocimiento, sexo, y convocatoria de admisión.
ST10	Personales, económicas, sociales, familiares y académicas.	La pedagogía de los profesores, sentimiento de frustración, importancia del programa, expectativas insatisfechas, motivación del programa y procrastinación.
ST11	Datos socioeconómicos, académicos y personales	Cursos aprobados, créditos aprobados, comunidad de origen, promedio acumulado y ciclo de ingreso.
ST12	Socioeconómica, familiar, personal y académica	Estado civil, tipo colegio, ciclo, ingresos familiares, cursos aprobados, vive con familia, ocupación padre, sexo, especialidad y turno.

Fuente. Elaboración propia

Otro aspecto importante se relaciona con la información presentada en la Tabla 3, donde se ilustran algunos métodos, herramientas, algoritmos y técnicas comúnmente empleadas.

Característica general	Detalle	ID del estudio
Metodología	Crisp-DM	ST1, ST11, ST12
	KDD	ST2, ST3, ST6, ST7, ST8
	Propia	ST4, ST5, ST9, ST10
Herramienta software	Rapid Miner	ST1, ST2, ST4
	Weka	ST3, ST7, ST8, ST10, ST11
	SPSS	ST9, ST12

Característica general	Detalle	ID del estudio
Herramienta software	Suite Pentaho	ST8
	Lenguaje R	ST5, ST6, ST9
	MySQL	ST1
	Oracle	ST8
Algoritmo de aprendizaje	Naive Bayes	ST1
	Árbol de decisión	ST1, ST2, ST3, ST4, ST5, ST7, ST8, ST9, ST10, ST11, ST12
	Árbol de clasificación y regresión	ST5
	Máquina de soporte vectorial	ST2
	Red neuronal artificial	ST2
	Regresión logística	ST2, ST6, ST8, ST9
	K vecinos más próximos	ST3
Selección de atributos	Reglas de inducción	ST3
	K-means/X-means	ST1, ST2
	GainRatioAttributeEval	ST3, ST7
	InfoGainAttributeEval	ST3, ST7, ST11
	Coefficiente de correlación de Pearson	ST10
Técnicas de Entrenamiento/Validación	Validación cruzada	ST1, ST8, ST10, ST11
	Balanceo	ST2
	Umbral de clasificación	ST2
	Matriz de confusión	ST1, ST4, ST6, ST7, ST8, ST10
	Curva ROC (Receiver Operating Characteristic)	ST1, ST5, ST8, ST11

Fuente. Elaboración propia

Población y muestra

La población objeto del estudio está constituida por estudiantes pertenecientes a la Universidad del Quindío en Colombia cuyo registro de matrícula corresponde al periodo comprendido entre el primer semestre de 2012 y el segundo semestre de 2018. Dichos estudiantes se encuentran agrupados en 7 facultades y 33 programas académicos de pregrado.

Se realizó un muestreo por conveniencia, formado por los casos disponibles a los cuales se tiene acceso (Hernández et al., 2014) y que fueron suministrados por la oficina de Admisiones y Registro y la oficina de Planeación. Una vez la información fue depurada se consolidó una muestra de 10705 estudiantes.

Procedimiento

El procedimiento aplicado se basa en el trabajo de Urbina-Nájera et al. (2020) en donde los autores proponen una adaptación de los pasos considerados por KDD. De esta forma se plantean las siguientes etapas: recolección de datos, pre-procesamiento de los datos, selección de atributos y aplicación de algoritmos de aprendizaje computacional. A continuación, se desarrollan cada una de estas etapas.

Etapas de recolección de datos. Los métodos de recolección se encargan de obtener y almacenar los registros proporcionados por las distintas fuentes de información y bases de datos existentes que poseen información relevante para la situación a tratar. En este trabajo, los datos se encontraron disponibles de forma digital, mediante hojas de cálculo en Excel proporcionadas por la oficina de Admisiones y Registro y la oficina de Planeación de la Universidad del Quindío.

La información proporcionada por la oficina de Admisiones y Registro estaba compuesta por un total de 672.996 registros y 22 atributos. En cuanto a la información provista por la oficina de Planeación se contó con un total de 201.569 registros y 39 atributos.

Etapas de pre-procesamiento de datos. Se refiere a refinar y filtrar los datos originales con el propósito de mejorar su procesamiento y accesibilidad. Esto permite la obtención de un conjunto de registros relevantes para el estudio, sin redundancia, que tengan completitud, que sean correctos y que no contengan información sensible (por ejemplo, nombre del estudiante o número de identificación personal), de acuerdo con las políticas gubernamentales establecidas para el tratamiento y protección de los datos.

Para lo lograr lo anterior se lleva a cabo unas tareas de filtrado, limpieza y transformación. Estas tareas se apoyaron en el uso de la herramienta Microsoft Excel® y el entorno de programación R Studio®

- a. Filtrado. En este punto se debe considerar un filtrado tanto a nivel atributos como de registros. Se procedió a descartar

la información sensible y los registros de estudiantes cuya matrícula no correspondía a un programa de estudio presencial.

- b. Limpieza. Se procedió a eliminar la mayor cantidad de registros erróneos, incompletos y redundantes. Además, se eliminaron algunos atributos considerados irrelevantes para el cumplimiento de los objetivos del proyecto. Algunos de estos atributos son: programa académico, tipo de documento, teléfono, nombre de acudiente, entre otros.
- c. Transformación. La actividad principal consistió en combinar e integrar los dos conjuntos de registros proporcionados por las dos oficinas de la Universidad del Quindío, los cuales tenían diferentes formatos, distinta cantidad de registros y estructuras heterogéneas. Por otra parte, se adicionaron columnas que permitieran la codificación de las características, ya que sus valores iniciales eran nominales, y dificultaba el tratamiento descriptivo de los mismos. Con el propósito de implementar algunas búsquedas y comparaciones, se procedió a concatenar los valores de algunas características como son nombres y apellidos, sin embargo; después de cumplir con su propósito inicial esta información sensible se eliminó. También se creó una columna para determinar si el estudiante se categoriza como desertor o no desertor. Para llevar a cabo esta categorización se aplicaron los criterios establecidos para dicha situación. Para la aplicación de estos criterios se debe revisar si un estudiante acumula dos periodos académicos (cada período académico corresponde a un semestre) seguidos sin matricularse y revisar durante toda la cohorte de análisis si a pesar de no matricularse durante dos periodos seguidos, un estudiante no se reintegra a la Universidad en cualquiera de sus facultades. Finalmente, el conjunto de toda la información se dividió en 7 conjuntos de datos que corresponden a cada una de las facultades de la Universidad: Agroindustria, Ciencias Básicas, Ciencias Económicas, Ciencias Humanas, Ciencias de la Salud, Ciencias de Educación e Ingeniería.

Como se puede apreciar, antes de ejecutar esta etapa se contaba inicialmente con 672.996 registros y 22 características, provistos por la oficina de Admisiones y Registro y con 203.308 registros y 39 características, provistos por la oficina de Planeación. Para un total de 876.304 registros y 61 características.

Al final de esta etapa se obtuvo un consolidado de la información distribuido en 7 conjuntos de datos (uno por cada facultad) y compuesto por 10.705 registros y por 13 características.

La correspondiente distribución, estructura y caracterización se ilustran en las Tablas 4, 5 y 6.

Tabla 4 Datos organizados e integrados por facultades

ID	Facultad	Cantidad de registros
FA	Agroindustria	374
FCB	Ciencias Básicas	1.261
FCE	Ciencias Económicas	2.051
FCH	Ciencias Humanas	1.128
FCS	Ciencias de la Salud	522
FE	Ciencias de Educación	2.906
FI	Ingeniería	2.463
	Total	10.705

Fuente. Elaboración propia

Tabla 5 Características organizadas e integradas

Características	Significado
GÉNERO	Tipo de género
ESTADO_CIVIL	Tipo de estado civil
VÍCTIMA_CONFLICTO_ARMADO	Si es o no víctima
ES_DESPLAZADO	Si es o no desplazado
ES_DISCAPACITADO	Si es o no discapacitado
ESTRATO_SOCIOECONÓMICO	Número de estrato
PROMEDIO_ACUMULADO	Promedio acumulado del estudiante
NÚCLEO_FAMILIAR	Conformación de la familia
APORTAN_EN_LA_FAMILIA	Si aportan o no
CANTIDAD_INGRESOS	Rango de los ingresos
LABORA	Si labora o no
MOTIVACIÓN	Motivo de ingreso a la universidad
DESERTOR	Estado desertor o no

Fuente. Elaboración propia

Tabla 6 Caracterización de la información

Característica	Valor	Asignación
GÉNERO	Femenino	0
	Masculino	1
ESTADO_CIVIL	Soltero(a)	0
	Casado(a)	1
	Separado(a)	2
	Unión libre	3
VÍCTIMA_CONFLICTO_ARMADO	No	0
	Si	1
ES_DESPLAZADO	No	0
	Si	1
ES_DISCAPACITADO	No	0
	Si	1
ESTRATO_SOCIOECONÓMICO	Entero entre [1 y 6]	1 - 6
PROMEDIO_ACUMULADO	Real entre [0 y 5]	0 - 5
NÚCLEO_FAMILIAR	Alguno de sus padres	0
	Dos padres y hermanos	1
	Esposa(o) o compañera y/o hijos	2
	Hermanos o familiares	3
	Otros	4
	Solo	5
	Solo sus dos padres	6
	Su(s) hijo (s)	7
APORTAN_EN_LA_FAMILIA	No	0
	Si	1
CANTIDAD_INGRESOS	Entre 0-1 salarios mínimos	1
	Entre 1-2 salarios mínimos	2
	Entre 2-5 salarios mínimos	3
	Entre 5-10 salarios mínimos	4
	Más de 10 salarios mínimos	5
LABORA	No	0
	Si	1

Característica	Valor	Asignación
MOTIVACIÓN	Por el prestigio de la educación universitaria	1
	Por la posibilidad de practicar los deportes que me gustan	2
	Porque me interesa formarme como investigado	3
	Porque me interesa la formación integral que ofrecen las universidades	4
	Porque puedo acceder después a postgrados	5
	Porque puedo conseguir un mejor empleo con la educación universitaria	6
DESERTOR	Desertor	0
	No desertor	1

Fuente. Elaboración propia

Etapa de selección de atributos. Cuando se tiene una gran cantidad de atributos que hacen parte de los datos recolectados, es conveniente ejecutar una tarea de selección de atributos con el propósito de mejorar la predicción, reducir tiempo de entrenamiento del algoritmo y reducir espacio de almacenamiento (Urbina-Nájera et al., 2020). Para llevar a cabo esta tarea se procede a determinar el grado de asociación entre las características, usando la prueba Chi-cuadrado. La prueba de Chi-cuadrado, se usa para determinar la existencia o ausencia de relaciones entre dos características categóricas. Es importante aclarar dicha prueba no indica el grado o el tipo de relación.

La prueba se aplicó a cada uno de los 7 conjuntos de datos que contienen la información de cada facultad. Se usó para determinar asociación de la característica aleatoria que contiene los resultados de los desertores y no desertores en cada facultad con cada una de las características perfiladas hasta el momento y que se ilustraron en la Tabla 5.

Para la selección de características se consideran los siguientes criterios:

1. Se excluyen las características que en una sola categoría presenten más de 80% de todos los casos, ya que pueden presentar un evidente desbalance entre los resultados.

2. Los que no cumplan con el criterio de asociación de la prueba Chi-cuadrado. Es decir, los resultados de la prueba Chi-cuadrado en donde p valor sea mayor que 5%, se excluyen.

Hipótesis:

H0: No existe asociación entre las características

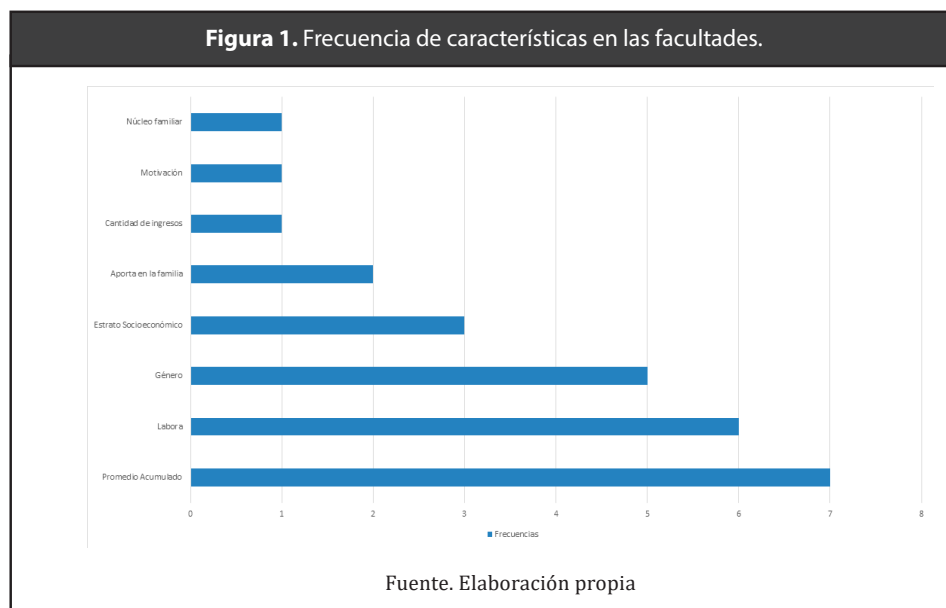
H1: Existe asociación entre las características

En la Tabla 7 se ilustran los resultados de las características seleccionadas para cada una de las facultades y en la Figura 1 se ilustra la frecuencia de ocurrencia de cada característica en las 7 facultades.

Tabla 7 Características seleccionadas por facultades

Facultad	Características
Agroindustria	Estrato socioeconómico, promedio acumulado
Ciencias Básicas	Género, cantidad de ingresos, labora, promedio acumulado
Ciencias Económicas	Género, labora, promedio acumulado
Ciencias Humanas	Género, labora, promedio acumulado
Ciencias de la Salud	Estrato socioeconómico, labora, aporta en la familia, promedio acumulado
Ciencias de Educación	Género, Estrato socioeconómico, labora, promedio acumulado
Ingeniería	Género, núcleo familiar, motivación, labora, aporta en la familia, promedio acumulado

Fuente. Elaboración propia



Etapa de aplicación de algoritmos de aprendizaje computacional. El aprendizaje computacional es el área de la Inteligencia Artificial que estudia los agentes o programas que aprenden y evolucionan a partir de su experiencia con el propósito de realizar una tarea cada vez mejor (Urbina-Nájera et al., 2020). La aplicación de algoritmos de aprendizaje computacional es la última etapa del proceso para el descubrimiento del conocimiento propuesto en este trabajo, la cual permite establecer algunas hipótesis a partir de la información que se tiene. Es así como se pueden obtener los modelos que permitan el análisis de la deserción estudiantil universitaria. Para lograr esto, es importante decidir sobre algunos aspectos que se relacionan con la calidad del conocimiento que se propone obtener y según Urbina-Nájera et al. (2020) son:

- *Identificar las tareas de minería de datos más apropiadas para el análisis.* Las cuales, según los autores, pueden ser descriptivas o predictivas. En este caso, se realizaron tareas descriptivas con el propósito de aumentar la comprensión de los datos disponibles, mediante la aplicación de medidas descriptivas tales como diagramas de barras y caja de bigotes. También se llevaron a cabo tareas predictivas con el propósito de generar modelos válidos para tratar futuros casos. Esto se determinó teniendo en cuenta los resultados de los estudios revisados y los objetivos propuestos en este trabajo.
- *Elegir el modelo teniendo en cuenta la forma en que se desea visualizar la información.* Con el propósito de ejecutar las tareas

predictivas se aplican modelos de clasificación, los cuales son los más utilizados en estos contextos educativos (Vila, 2019). La aplicación de estos modelos se basa en la elaboración de un modelo a partir de unos datos de entrenamiento. Dicho modelo permitirá clasificar individuos con base en una característica cualitativa. Para realizar la clasificación se consideró el uso de árboles de decisión, ya que están provistos por un modelo simple que se puede interpretar con bastante facilidad y que además ofrecen una velocidad de clasificación eficiente (Vila, 2019). El aprendizaje basado en árboles de decisión utiliza un árbol de decisión como un modelo predictivo que mapea observaciones sobre un elemento a conclusiones sobre el valor objetivo del elemento.

- *Elegir el algoritmo más eficiente para resolver la tarea y devolver el modelo deseado.* En el contexto de la minería de datos se encuentran diversos algoritmos de árboles de decisión. En este caso, se usó el algoritmo árboles de decisión provisto por la herramienta RapidMiner© basado en C4.5. Esto se determinó teniendo en cuenta los estudios revisados que evidencian el repetido uso de este algoritmo en la solución de situaciones educativas en el nivel superior y con base en algunos estudios, como Yamao et al. (2018), quienes reportan que este algoritmo en comparación con otros similares demuestra un mayor porcentaje de predicción. También, según Urbina-Nájera et al. (2020), este algoritmo proporciona una mejor comprensión del conocimiento y ofrece otras ventajas importantes tales como la reducción del número características independientes y la representación visual de un problema, entre otras. De igual forma, otros autores afirman que dicho algoritmo se utiliza con frecuencia en la identificación de características relevantes dentro un conjunto de datos con una menor tasa de error y mayor precisión, entre otras ventajas asociadas a la toma de decisiones (Gupta et al., 2017; Sharma & Kumar, 2016).

3. Resultados

En esta sección se presentan los resultados de la aplicación del proceso para el descubrimiento de conocimiento en función de los análisis descriptivos y predictivos mencionados en la sección anterior.

En cuanto al análisis descriptivo, se presentan en las Figuras 2a, 2b y 2c las distribuciones en cada facultad de los promedios

acumulados de los estudiantes en función de la característica dependiente que determina el estado desertor o no. Se decidió presentar en este artículo dicha distribución, considerando que la característica *Promedio Acumulado* fue la única característica que el proceso de selección de características determinó en común para todas las facultades y además es cuantitativa. Este análisis se hace mediante la aplicación de diagramas de cajas y bigotes revisando las medidas que se generan. Estos diagramas son de gran utilidad ya que permiten representar de forma gráfica la distribución de puntuaciones dentro de una característica. También permiten señalar los valores atípicos o casos extremos de la característica. En este análisis, se considera la distribución de los datos de desertores y no desertores, con respecto al promedio acumulado. Se encuentra que, para todas las facultades, la dispersión entre los datos de los estudiantes desertores es mayor que la de los no desertores; es decir, existe mayor homogeneidad en los promedios acumulados de los estudiantes desertores. Dicha homogeneidad se refiere a que los promedios de los estudiantes no desertores son más parecidos. Por otra parte, se encuentra un mayor número de valores atípicos en los estudiantes desertores y sus promedios se encuentran por debajo de rangos que varían dependiendo de la facultad, siendo estos: *Agroindustria* (1.9875), *Ciencias Básicas*, (1.5), *Ciencias Económicas*, (1.73), *Ciencias Humanas* (1.95625), *Ciencias de la Salud* (0.6825), *Ciencias de la Educación* (1.655) e *Ingeniería* (0.975).

En cuanto al análisis predictivo y partiendo de las características perfiladas en la Tabla 6, se llevó a cabo un proceso de identificación de las características más relevantes aplicando evaluadores de selección de atributos. Una vez aplicado el evaluador Chi cuadrado se obtuvo un subconjunto de 8 principales características extraídas del conjunto de datos (organizados e integrados) que contenía 18 características. Es importante recordar que el conjunto de datos inicial con el que se trabajó y antes de realizar cualquier depuración, estaba compuesto por 61 características. Posteriormente, se aplicó el procedimiento para reconocimiento de patrones de la deserción universitaria con árboles de decisión, mediante RapidMiner© con la funcionalidad Decision Tree y el operador Optimize Parameters, que permite encontrar los valores óptimos de los parámetros.

Figura 2a. Distribución respecto al promedio acumulado en cada facultad.

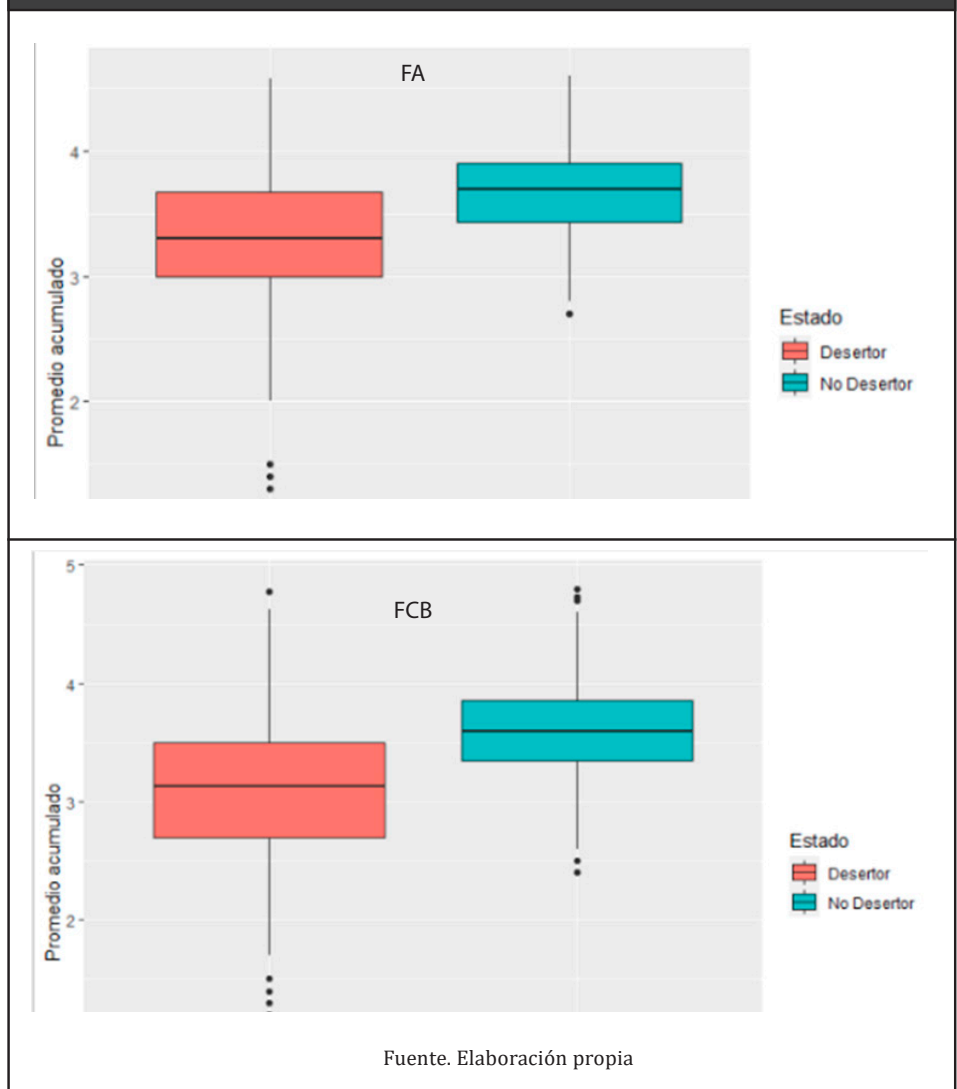


Figura 2b. Distribución respecto al promedio acumulado en cada facultad.

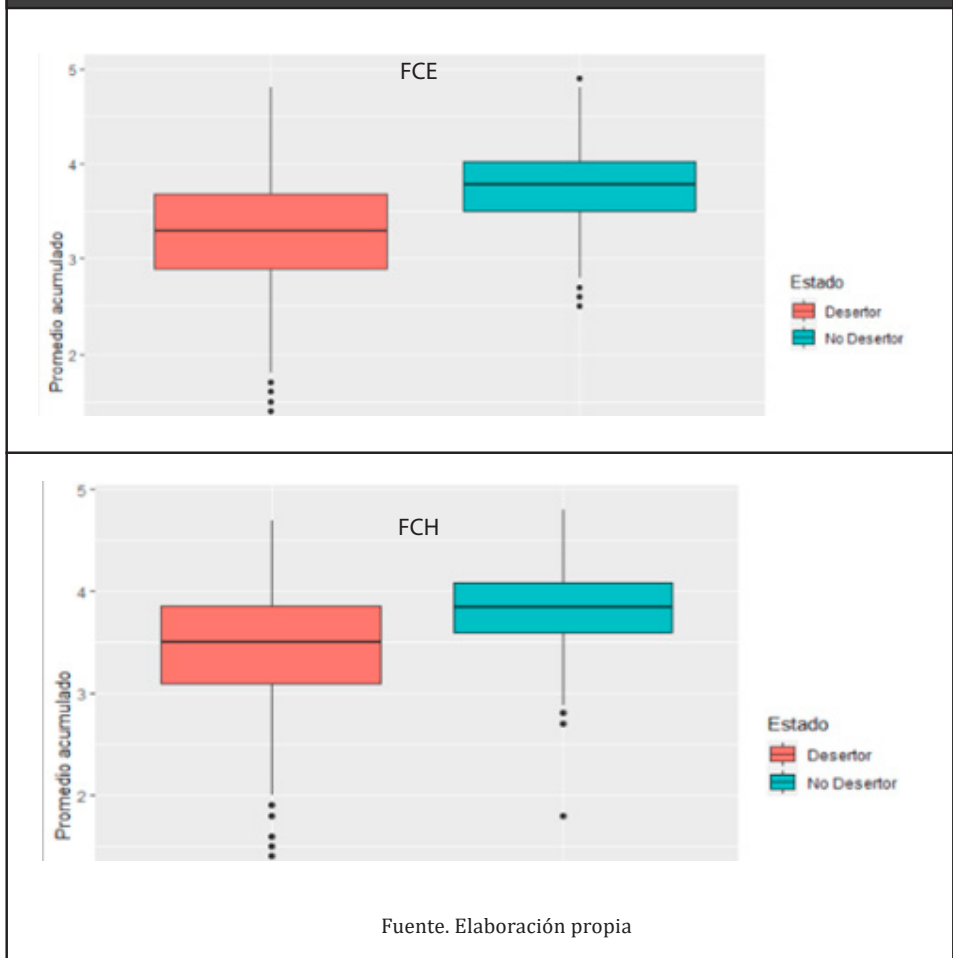
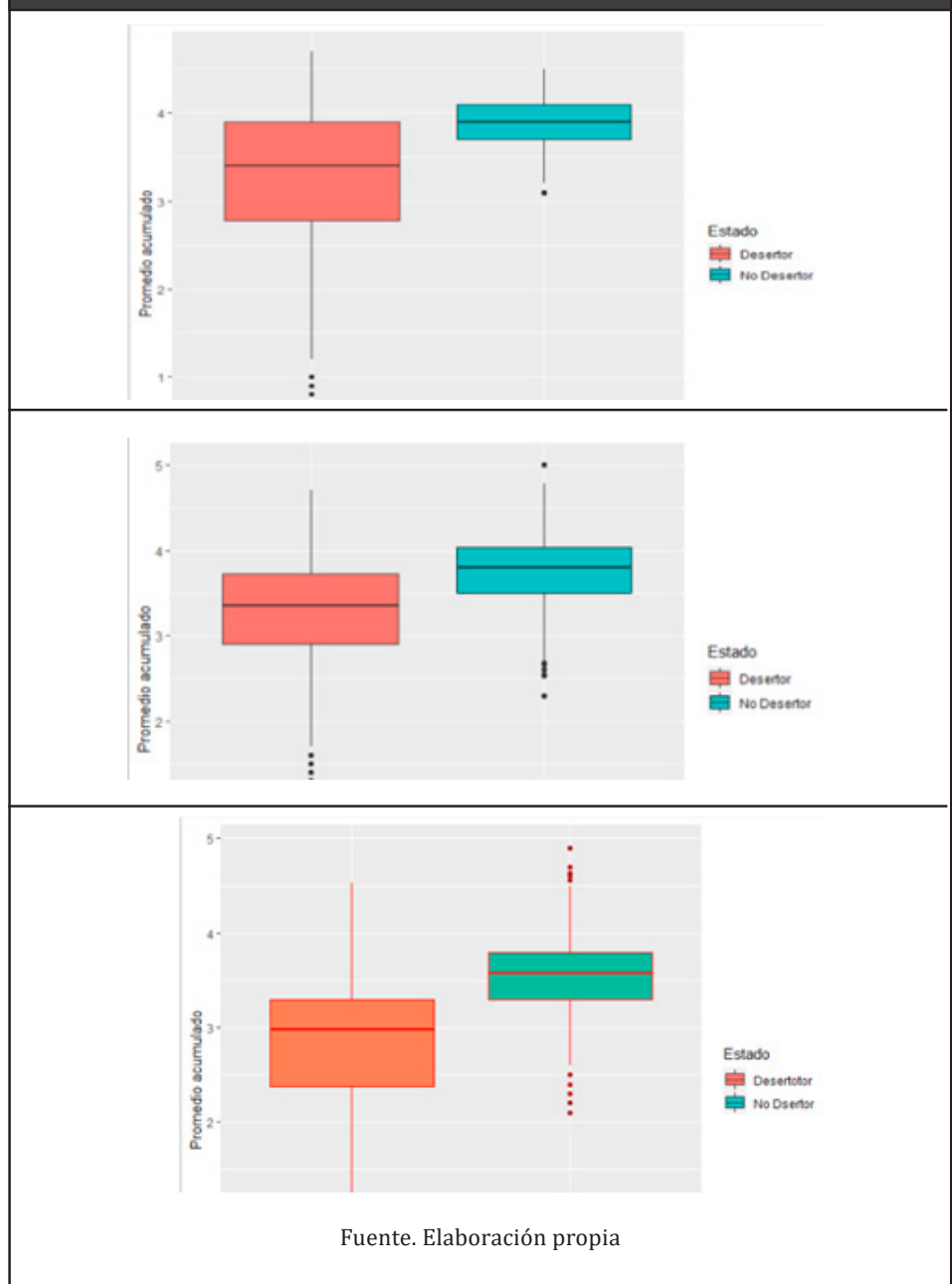


Figura 2c. Distribución respecto al promedio acumulado en cada facultad.



Luego de obtener los modelos de minería de datos, se lleva a cabo la evaluación de la calidad de dichos modelos y su posterior interpretación. Para la aplicación del algoritmo de clasificación, cada conjunto de datos se dividió en un 70% para entrenamiento y un 30% para validación. Para evaluar los algoritmos aplicados se emplearon métricas de calidad usando la matriz de confusión y verificando la certeza en la predicción y rendimiento del modelo de clasificación. La matriz de confusión, tal como se muestra en la

Figura 3, es una matriz cuadrada que consta de dos características principales positivo o negativo tanto en las filas como en las columnas. Los valores de las columnas representan los valores actuales y los de las filas los valores predichos.

Figura 3. Matriz de confusión.

		Observación	
		Positivo (1)	Negativo (0)
Valores predictivos	Positivo (1)	VP	FP
	Negativo (0)	FN	VN

Fuente. Elaboración propia

La definición de la matriz que se ilustra en la Figura 3 y con relación al problema de investigación es:

VP: El estudiante es desertor y el modelo lo predijo como desertor

FP: El modelo lo predijo como desertor, pero es no desertor

FN: El modelo lo predijo como no desertor, pero es desertor

VN: El estudiante es no desertor y el modelo lo predijo como no desertor

A continuación, se describen algunas métricas empleadas para evaluar el modelo, cuyos resultados se ilustran en la Tabla 8.

1. *Recall (sensibilidad)*. Es la proporción de positivos predichos por el modelo que son realmente verdaderos positivos.
2. *Accuracy (Exactitud)*. Es la proporción de predicciones correctas hechas por el modelo con respecto al número de registros.
3. *Precision (Precisión)*. Es la proporción de positivos que son realmente verdaderos positivos.
4. *Balanced Accuracy*. Es el promedio de las proporciones correctas de cada clase individualmente.

Los resultados generados en la matriz de confusión por cada una de las facultades se presentan en la tabla 7 y de acuerdo con los valores de las métricas de la Tabla 8 se puede determinar que el rendimiento de los modelos generados es satisfactorio.

Tabla 7 Matrices de confusión

Facultad	Matrices de Confusión		
		True 0	True 1
Ciencias Agroindustriales			
	Pred 1	14	83
	Pred 0	13	2
Ciencias Básicas y Tecnología		True 0	True 1
	Pred 1	73	204
	Pred 0	88	14
Ciencias Económicas y Administrativas		True 0	True 1
	Pred 1	91	436
	Pred 0	70	18
Ciencias Humanas y Bellas Artes		True 0	True 1
	Pred 1	49	254
	Pred 0	31	4
Ciencias de la Salud		True 0	True 1
	Pred 1	14	106
	Pred 0	33	4
Ciencias de la Educación		True 0	True 1
	Pred 1	183	513
	Pred 0	145	31
Ingeniería		True 0	True 1
	Pred 1	42	228
	Pred 0	362	107

Tabla 8 Indicadores para validación de los modelos

Facultad	Recall	Accuracy	Precision	Balanced Accuracy
FA	0.98	0.86	0.86	0.73
FCB	0.94	0.77	0.74	0.74
FCE	0.96	0.82	0.83	0.70
FCH	0.98	0.84	0.84	0.69
FCS	0.96	0.89	0.88	0.83
FE	0.94	0.75	0.74	0.69
FI	0.68	0.80	0.84	0.79

Fuente. Elaboración propia

Una vez efectuado el análisis de los resultados de cada facultad, se descubrieron algunos patrones que se ilustran en las Figuras 4 y 5. Para interpretar los datos se procedió a recorrer los resultados desde la raíz hacia las hojas cuyo atributo de clase sea DESERTOR. En este caso el valor = 0 indica que un estudiante es desertor y el valor = 1 indica que un estudiante no es desertor.

Figura 4. Patrones facultad de Agroindustria.

Facultad Agroindustria

```

PROMEDIO_ACUMULADO > 2.640
--PROMEDIO_ACUMULADO > 3.040
----PROMEDIO_ACUMULADO > 3.215
-----PROMEDIO_ACUMULADO > 3.295
-----PROMEDIO_ACUMULADO > 3.315:1 (No Desertor)
-----PROMEDIO_ACUMULADO <= 3.315
-----ESTRATO_SOCIOECONÓMICO = 1:1 (No Desertor)
-----ESTRATO_SOCIOECONÓMICO = 2:1 (No Desertor)
-----ESTRATO_SOCIOECONÓMICO = 3 0 (Desertor)
-----PROMEDIO_ACUMULADO <= 3.295:1 (No Desertor)
----PROMEDIO_ACUMULADO <= 3.215
-----PROMEDIO_ACUMULADO > 3.075
-----PROMEDIO_ACUMULADO > 3.195
-----ESTRATO_SOCIOECONÓMICO = 1:1 (No Desertor)
-----ESTRATO_SOCIOECONÓMICO = 2:0 (Desertor)
-----ESTRATO_SOCIOECONÓMICO = 3:1 (No Desertor)
-----PROMEDIO_ACUMULADO <= 3.195
-----PROMEDIO_ACUMULADO > 3.175:1 (No Desertor)
-----PROMEDIO_ACUMULADO <= 3.175
-----PROMEDIO_ACUMULADO > 3.145:1 (No Desertor)
-----PROMEDIO_ACUMULADO <= 3.145:0 (Desertor)
-----PROMEDIO_ACUMULADO <= 3.075:1 (No Desertor)
--PROMEDIO_ACUMULADO <= 3.040
----PROMEDIO_ACUMULADO > 3.015:0 (Desertor)
----PROMEDIO_ACUMULADO <= 3.015:1 (No Desertor)
PROMEDIO_ACUMULADO <= 2.640:0 (Desertor)

```

Fuente. Elaboración propia

Figura 5. Patrones otras facultades.

Figura 5. Patrones otras facultades.	
<p>Facultad Ciencias Básicas</p> <p>PROMEDIO_ACUMULADO > 3.085:1 PROMEDIO_ACUMULADO <= 3.085:1</p>	<p>Facultad Ciencias Económicas</p> <p>PROMEDIO_ACUMULADO > 2.450 --PROMEDIO_ACUMULADO > 2.805 ----PROMEDIO_ACUMULADO > 2.955 -----PROMEDIO_ACUMULADO > 3.185:1 -----PROMEDIO_ACUMULADO <= 3.185 -----GÉNERO = 0:0 -----GÉNERO = 1:1 ---PROMEDIO_ACUMULADO <= 2.955:0 --PROMEDIO_ACUMULADO <= 2.805:0 PROMEDIO_ACUMULADO <= 2.450:0</p>
<p>Facultad Ciencias Humanas</p> <p>PROMEDIO_ACUMULADO > 1.700 --PROMEDIO_ACUMULADO > 2.805 ----PROMEDIO_ACUMULADO > 3.005 -----PROMEDIO_ACUMULADO > 3.255:1 -----PROMEDIO_ACUMULADO <= 3.255 -----GÉNERO = 0: 1 -----GÉNERO = 1: 0 ---PROMEDIO_ACUMULADO <= 3.005:0 --PROMEDIO_ACUMULADO <= 2.805:0 PROMEDIO_ACUMULADO <= 1.700:0</p>	<p>Facultad Ciencias de la Salud</p> <p>PROMEDIO_ACUMULADO > 3.205 --PROMEDIO_ACUMULADO > 3.505 ----PROMEDIO_ACUMULADO > 4.290 -----APORTAN_EN_LA_FAMILIA = 0:0 -----APORTAN_EN_LA_FAMILIA = 1:1 ---PROMEDIO_ACUMULADO <= 4.290:1 --PROMEDIO_ACUMULADO <= 3.505 ---APORTAN_EN_LA_FAMILIA = 0:0 ---APORTAN_EN_LA_FAMILIA = 1:1 PROMEDIO_ACUMULADO <= 3.205:0</p>
<p>Facultad Ciencias de Educación</p> <p>PROMEDIO_ACUMULADO > 3.305:1 PROMEDIO_ACUMULADO <= 3.305 --PROMEDIO_ACUMULADO > 2.515 ----PROMEDIO_ACUMULADO > 2.955 -----LABORA = 0 -----PROMEDIO_ACUMULADO > 3.245:1 -----PROMEDIO_ACUMULADO <= 3.245 -----GÉNERO = 0:1 -----GÉNERO = 1:0 -----LABORA = 1:0 ---PROMEDIO_ACUMULADO <= 2.955:0 --PROMEDIO_ACUMULADO <= 2.515:0</p>	<p>Facultad de Ingeniería</p> <p>PROMEDIO_ACUMULADO > 3.085 --PROMEDIO_ACUMULADO > 3.335:1 --PROMEDIO_ACUMULADO <= 3.335 ---LABORA = 0 -----APORTAN_EN_LA_FAMILIA = 0:1 -----APORTAN_EN_LA_FAMILIA = 1 -----PROMEDIO_ACUMULADO > 3.275:1 -----PROMEDIO_ACUMULADO <= 3.275 -----GÉNERO = 0:0 -----GÉNERO = 1:1 ---LABORA = 1:0</p>
<p>Fuente. Elaboración propia</p>	

4. Discusión y Conclusiones

Como resultado de la aplicación de la minería de datos educativa en el contexto del descubrimiento de conocimiento en bases de datos, se puede concluir que las principales características que inciden en la deserción estudiantil, en el contexto de toda la universidad, corresponden a núcleo familiar, motivación, cantidad de ingresos, aporta en la familia, estrato socioeconómico, género, labora y promedio acumulado. Siendo las 5 últimas las que presentan mayores frecuencias de incidencias en las distintas facultades y teniendo a promedio acumulado como la única característica en común para las 7 facultades estudiadas. Estos hallazgos coinciden con los resultados que se encontraron en la mayoría de los estudios revisados (ST1, ST2, ST3, ST4, ST5, ST6, ST8, ST11), para los cuales, las características relacionadas con el desempeño académico tienen una alta influencia en la deserción. Por otra parte, difiere de algunos estudios (ST7, ST9, ST10, ST12) que no consideran el desempeño académico como una característica relevante y en cambio proponen otras características como las más influyentes, algunas de ellas también fueron analizadas en este estudio, pero otras no se incluyeron por no contar con dicha información. Esta situación permite inferir que el contexto influye de manera determinante en el desarrollo y resultados de este tipo de estudios; además, permite reconocer la oportunidad y pertinencia de recolectar información adicional que hasta el momento no se había considerado. Lo anterior, con el propósito de generar estudios cada vez más acertados que apoyen y mejoren continuamente los procesos de toma de decisiones. Por otro lado, mediante la aplicación del algoritmo basado en árboles de decisión se logró identificar un conjunto de patrones que conllevan a una deserción universitaria.

Finalmente, se puede concluir que la deserción estudiantil universitaria no depende de una sola característica, sino que es causada por un conjunto de características y su interrelación, tal como se evidencia en los patrones identificados.

5. Agradecimientos

Este artículo se elaboró en el marco del proyecto de investigación No. 914 de la Universidad del Quindío titulado “Análisis de la deserción estudiantil en la Universidad del Quindío aplicando procesos de descubrimiento de conocimiento sobre datos”, el cual contó con el apoyo y patrocinio de la Vicerrectoría de Investigaciones de la Universidad del Quindío, Colombia.

6. Referencias bibliográficas

- Ayala, E., López, R. & Menéndez, V. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *Revista de Educación a Distancia (RED)*, 21(66), 1-36. <https://doi.org/10.6018/red.463561>
- Bakhshinategh, B., Zaiane, O. R., Elatia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: a survey of the last 10 years. *Education and Information Technologies*, 23(1), 537-553. <https://doi.org/10.1007/s10639-017-9616-z>
- Castro, L. F., Espitia, E. & Cardona, S. (2019). Analysis of Student Desertion in a Systems and Computing Engineering Undergraduate Program. *Revista Colombiana de Computación*, 20(1), 72-82. <https://doi.org/10.29375/25392115.3608>
- Castro, L. F., Espitia, E. & Mantilla, A. (2018). Applying CRISP-DM in a KDD Process for the Analysis of Student Attrition. *Communications in Computer and Information Science*, 885, 386-401. https://doi.org/10.1007/978-3-319-98998-3_30
- Castrillón-Gómez, O. D., Sarache W., & Ruiz-Herrera, S. (2020). Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formación Universitaria*, 13(6), 217-228. <http://dx.doi.org/10.4067/S0718-50062020000600217>
- Clerici, R., & Da Re, L. (2019). Evaluación de la eficacia de un programa de tutoría formativa. *Revista de Investigación Educativa*, 37(1), 39-56. <http://dx.doi.org/10.6018/rie.37.1.322331>
- Constante, A., Florenciano, E., Navarro, E. & Fernández, M. (2021). Factores asociados al abandono universitario. *Educación XX1*, 24(1), 17-44. <http://doi.org/10.5944/educXX1.26889>
- Cuji, B., Gavilanes, W., & Sánchez, R. (2017). Modelo predictivo de deserción estudiantil basado en arboles de decisión. *Espacios*, 38(55), 19-25. <https://www.revistaespacios.com/a17v38n55/a17v38n55p17.pdf>
- Ghazal, M. & Hammad, A. (2022) Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects. *International Journal of Construction Management*, 22(9), 1632-1646. <https://doi.org/10.1080/15623599.2020.1738205>.
- Gupta, B., Rawat, A., Jain, A., Arora, A., & Dhama, N. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications*, 163(8), 15-19. <https://doi.org/10.5120/ijca2017913660>
- Hatos, A., Coloja, R. & Sava, A. (2020). Assessing Situational Awareness of Universities Concerning Student Dropout: A Web-Based Content Analysis of Romanian Universities' Agenda. *Journal of Research in Higher Education*, 4 (2), 18-34. <https://doi.org/10.24193/JRHE.2020.2.2>
- Hernández, R., Fernández, C., & Baptista, M. (2014). *Metodología de la Investigación*. McGraw Hill Educación.

- Kumar, M., Singh, A. J., & Handa, D. (2017). Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques. *International Journal of Education and Management Engineering*, 6, 40-49. <https://doi.org/10.5815/ijeme.2017.06.05>
- Ministerio de Educación Nacional. (2021). Estadísticas de deserción y permanencia en educación superior, históricos indicadores 2010-2018. https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-357549_recurso_7.pdf
- Munizaga, F., Cifuentes, M., & Beltrán, A. (2018). Retención y abandono estudiantil en la Educación Superior Universitaria en América Latina y el Caribe: Una revisión sistemática. *Archivos Analíticos de Políticas Educativas*, 26(61), 1-36. <http://dx.doi.org/10.14507/epaa.26.3348>
- Oficina Europea de Estadística. (2020). Early leavers from education and training. [Mensaje en un blog]. Blog Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php/Early_leavers_from_education_and_training#Overview
- Oñate, A. A. (2016). *Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería de datos*. [Tesis de Maestría, Universidad Nacional de Colombia]. Repositorio Universidad Nacional. <https://repositorio.unal.edu.co/handle/unal/57387>
- Pando, A. & Zarate, W. (2020). *Aplicación de un modelo de minería de datos para identificación de patrones que influyen en la deserción académica en el instituto superior Leonardo Davinci*. [Trabajo de grado, Universidad Privada Antenor Orrego]. Repositorio de Tesis UPAO. <https://hdl.handle.net/20.500.12759/7033>
- Proyecto ALFA-GUIA. (2013). *Marco Conceptual sobre el Abandono*. https://documentop.com/marco-conceptual-abandono-proyecto-alfa-guia_59fbf0b21723dda8a11794fa.html
- Quiñones, L., Jara, D., Alvarado, N., Milla, M. & Gamarra, O. (2020). Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *RECyT*, 34, 45-50. <https://doi.org/10.36995/j.recyt.2020.34.006>
- Ramírez, P., & Grandón, E. (2018). Predicción de la deserción académica en una universidad pública chilena a través de la clasificación basada en árboles de decisión con parámetros optimizados. *Formación Universitaria*, 11(3), 3-10. <http://dx.doi.org/10.4067/S0718-50062018000300003>
- Ramírez, V. (2021). *Deserción estudiantil y el costo económico en universidades chilenas*. [Tesis de Maestría, Universidad del Bio-Bio]. Repositorio digital Universidad del Bio-Bio. <http://repobib.ubiobio.cl/jspui/handle/123456789/3609>
- Sharma, H., & Kumar, S. (2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *International Journal of Science and Research*, 5 (4), 2094-2097. <https://doi.org/10.21275/v5i4.NOV162954>
- Urbina-Nájera, A. B., Camino-Hampshire, J. C., & Cruz-Barbosa, R. (2020). Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *RELIEVE*, 26(1), 1-21. <http://doi.org/10.7203/relieve.26.1.16061>
- Vásquez, J. (2016). *Modelo predictivo para estimar la deserción de estudiantes en una Institución de Educación Superior*. [Tesis de Maestría, Universidad de Chile]. Repositorio Académico de la Universidad de Chile. <http://repositorio.uchile.cl/handle/2250/144169>
- Vicente, V. X. (2020). *Aplicación de la técnica de minería de datos para la predicción de la deserción estudiantil universitaria*. [Trabajo de grado, Universidad Técnica de Ambato]. Repositorio Universidad Técnica de Ambato. <https://repositorio.uta.edu.ec/jspui/bitstream/123456789/30892/1/Victor%20Xavier%20Vicente%20Guerrero..pdf>
- Vila, D. (2019). *Detección de patrones de deserción estudiantil utilizando técnicas predictivas de clasificación y regresión de minería de datos*. [Trabajo de grado, Universidad Técnica del Norte]. Repositorio Digital Universidad Técnica del Norte. <http://repositorio.utn.edu.ec/handle/123456789/9095>
- Villalobos, L. R. (2017). *Enfoques y diseños de investigación social: cuantitativos, cualitativos y mixtos*. EUNED.

- Villanueva, A. (2016). *Análisis de deserción en una Institución de Educación Superior Colombiana mediante el uso de Minería de Datos para la Educación (EDM)*. [Tesis de Maestría, Institución universitaria Politécnico Gran Colombiano]. Repositorio SISNAB. <http://hdl.handle.net/10823/943>
- Yamao, E., Saavedra, L. C., Campos, R., & Huancas, V. D. J. (2018). Prediction of academic performance using data mining in first year students of peruvian university. *CAMPUS*, 13(26), 151-160. <https://doi.org/10.24265/campus.2018.v23n26.05>
- Zarate, A. J. (2019). *Reconocimiento de patrones de deserción de estudiantes universitarios basados en modelos de clasificación*. [Trabajo de grado, Universidad Nacional de San Agustín de Arequipa]. Repositorio institucional UNSA. <http://repositorio.unsa.edu.pe/handle/UNSA/9419>