

# 8. PROPUESTA DE APLICACIÓN DE MINERÍA DE DATOS EN EL CONTEXTO EDUCATIVO UNIVERSITARIO: ESTUDIO DE CASO

## Proposal for the Application of Data Mining in the University Educational Context: A Case Study

*Luis Fernando Castro Rojas <sup>1</sup>, Esperanza Espitia Peña <sup>2</sup>*

***Fecha recibido:*** 23/11/2022

***Fecha aprobado:*** 15/12/2022

***Derivado del proyecto:*** “Análisis del rendimiento académico de los estudiantes de la Universidad del Quindío aplicando técnicas de minería de datos (Proyecto de investigación nro. 1083)”.

***Institución financiadora:*** Universidad del Quindío

***Pares evaluadores:*** Red de Investigación en Educación, Empresa y Sociedad – REDIEES.

---

<sup>1</sup> Ingeniero de sistemas, PhD, docente, Universidad del Quindío, [lufer@uniquindio.edu.co](mailto:lufer@uniquindio.edu.co).

<sup>2</sup> Ingeniera de Sistemas, MSc, docente, Universidad del Quindío, [eespitia@uniquindio.edu.co](mailto:eespitia@uniquindio.edu.co).

## RESUMEN

Las instituciones de educación superior enfrentan retos cada vez más exigentes para garantizar el desarrollo exitoso de su misión. Una preocupación frecuente de dichas instituciones es el fracaso académico de los estudiantes como consecuencia de un mal rendimiento académico. Esto ocasiona dificultades tanto para el estudiante como para la institución, que se materializan en pérdidas considerables de dinero y tiempo. Estas pérdidas se pueden deber a retrasos en el plan de estudios, reprobación de asignaturas, abandono de la carrera profesional, reinicio en los procesos de formación, deserción estudiantil o desaprovechamiento de recursos, entre otros. Este problema es complejo debido a la incidencia de múltiples factores, lo cual se evidencia en la literatura, que cuestiona la falta de soluciones definitivas para superar esta situación. Como aporte a la solución de dicho problema, en este trabajo se propone una metodología basada en técnicas de minería de datos educativos conocidas como EDM (Educational Data Mining), con el propósito de entender y analizar el fenómeno del rendimiento académico. Además, se presentan hallazgos en el contexto de la problemática mencionada obtenidos a partir de revisión de literatura y, finalmente, se presentan los resultados de las etapas iniciales de la aplicación de dicha metodología a un estudio de caso real en una universidad pública colombiana. Lo anterior permite concluir acerca de la pertinencia de la aplicación de MDE en el contexto del desempeño académico de los estudiantes universitarios.

**PALABRAS CLAVE:** *educación superior; rendimiento académico; MDE; KDD; Minería de datos.*

## **ABSTRACT**

Higher education institutions face increasingly demanding challenges to guarantee the successful development of their mission. A frequent concern of these institutions is the academic failure of students as a consequence of poor academic performance. This causes difficulties for both the student and the institution, which materialize in considerable losses of money and time. These losses may be due to delays in the study plan, failure of courses, abandonment of the professional career, restarting the training processes, student attrition, or waste of resources, among others. This problem is complex due to the incidence of multiple factors, which is evidenced in the literature, which questions the lack of definitive solutions to overcome this situation. As a contribution to solving this problem, this paper proposes a methodology based on educational data mining techniques known as EDM (Educational Data Mining), with the purpose of understanding and analyzing the phenomenon of academic performance. In addition, some findings are presented in the context of the mentioned problem, which are obtained from literature review. Finally, the results of the initial stages of the application of said methodology to a real case study in a Colombian public university are presented. The foregoing allows us to conclude about the relevance of the MDE application in the context of the academic performance of university students.

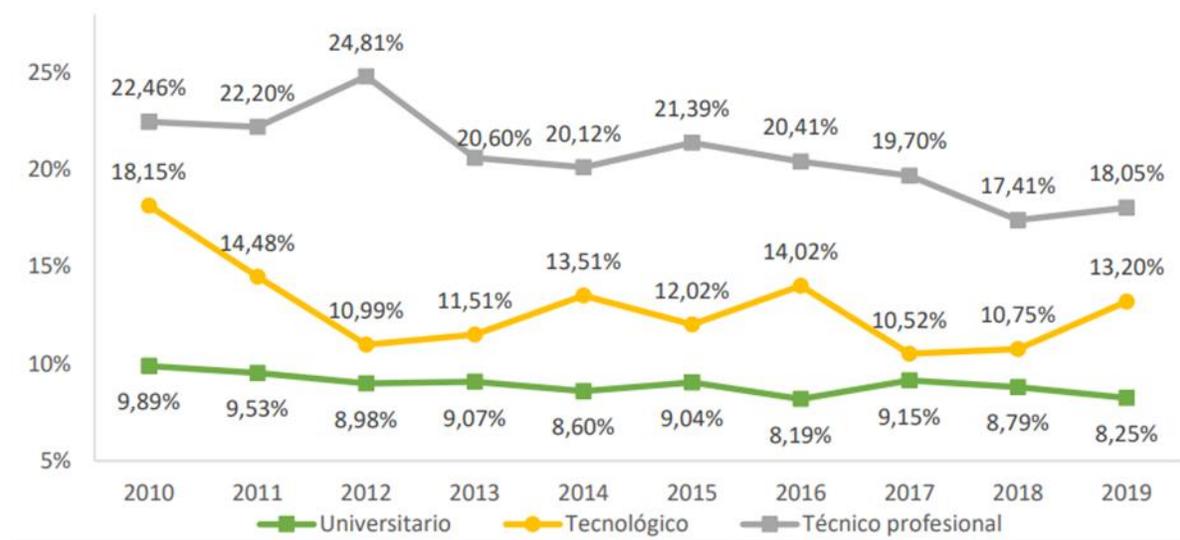
**KEYWORDS:** *higher education; academic performance; MDE; KDD; data mining.*

## INTRODUCCIÓN

Las instituciones de educación superior tienen algunos retos que deben cumplir para lograr el reconocimiento y posicionamiento en cuanto a la calidad, logrando niveles de éxito y satisfacción en los usuarios. En este caso, la propuesta que se presenta aborda de forma directa la necesidad de mantener y fortalecer cada vez más los mecanismos y estrategias de apoyo en los estudiantes de la universidad del Quindío orientadas a alcanzar un buen rendimiento académico mediante el uso de técnicas que propicien mejoras en las diferentes actividades realizadas en los espacios académicos registrados y en general en las políticas y estrategias que la institución genere en torno al fenómeno del rendimiento académico estudiantil. Es en este sentido, que se refleja la necesidad de analizar algunos factores que influyen en el rendimiento académico, los cuales no solo inciden en los estudiantes evitando pérdidas de dinero y de tiempo ocasionados por las diversas interrupciones, abandonos, inicios de nuevas carreras, entre otros, a los que el estudiante se ve sometido como consecuencia de un bajo rendimiento académico. Si no que, de manera directa, la institución también se va a ver favorecida en el momento en que sea evaluada por entidades, dependencias, sector industrial y empresarial y por el ministerio de educación para observar y posicionar su actividad académica en un buen nivel; así como, mediante la optimización de costos, tiempos y recurso humano demandados por los trámites administrativos desgastantes que requiere el manejo de las situaciones asociadas al bajo rendimiento académico. Para sustentar lo anterior se presentan en las Figuras 1 y 2, algunos de los indicadores de calidad establecidos por el Ministerio de Educación colombiano y que son influenciados directamente por el desempeño académico de los estudiantes.

**Figura 1.**

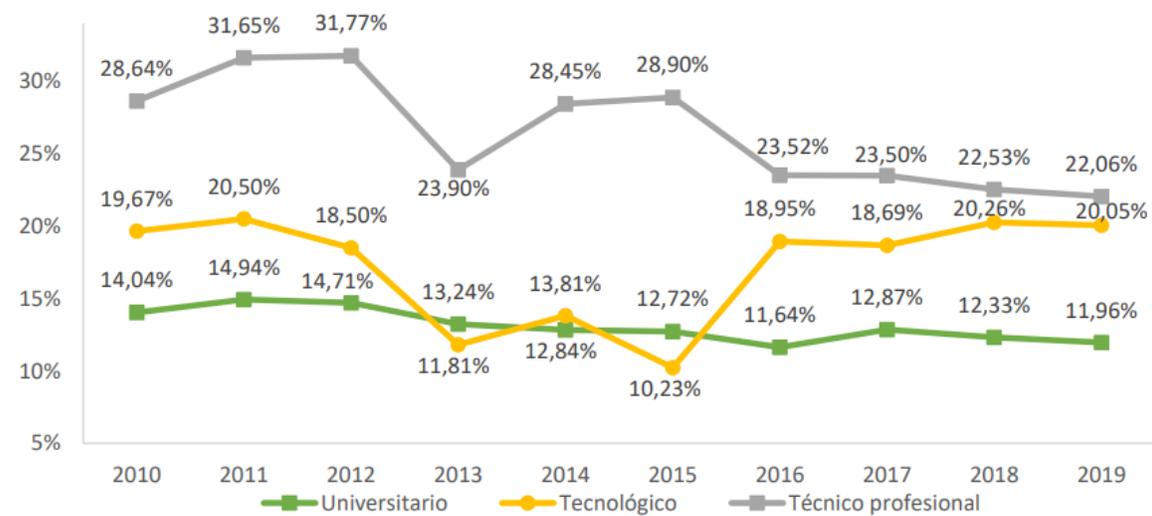
*Tasa de deserción anual según nivel de formación*



*Nota.* SPADIES Corte de los datos: noviembre de 2020.

**Figura 2.**

*Tasa de ausencia intersemestral\* según nivel de formación*



*Nota.* SPADIES Corte de los datos: noviembre de 2020. \*La tasa de ausencia intersemestral (TAI), se refiere a la proporción de estudiantes que se encuentran en riesgo de desertar al no matricularse en un programa académico durante un semestre.

Esta propuesta se basa en la capacidad de predecir el rendimiento académico de los estudiantes, lo cual es algo crucial para cualquier institución educativa, ya que permite obtener información relevante para la creación de nuevas reglas y estándares que apoyen un mejoramiento de su educación y su reputación (Kumar, *et al.*, 2017). Según Menacho (2017), el rendimiento académico de los estudiantes es uno de los temas de mayor preocupación que deben abordar las instituciones educativas superiores y en tal sentido, las técnicas de minería de datos (TMD) aplicadas a los datos generados en los ambientes educativos, están demostrando ser herramientas eficaces para predecir el rendimiento académico de los estudiantes; con la finalidad de identificar los factores que más influyen en su aprendizaje y apoyar a los profesores a mejorar el proceso de enseñanza a través de realizar acciones pedagógicas más eficientes y oportunas.

Adicionalmente, tal como se contempla en esta propuesta, el uso de la disciplina EDM en el contexto educativo de la Universidad del Quindío tiene el potencial de transformar los modelos existentes de enseñanza-aprendizaje al proporcionar nuevas herramientas de análisis, interacción e intervención (Aldowah, Al-Samarraie y Fauzy, 2019). La minería de datos educativa ayuda a encontrar información oculta a partir de las bases de datos existentes en la Universidad del Quindío, las cuales proveen una gran cantidad de datos e información de la institución educativa relacionados con el estudiante.

Por otra parte, con respecto ámbito investigativo-científico, la disciplina que se pretende adoptar en esta propuesta, cuenta con métodos que se pueden utilizar para el desarrollo de investigaciones aplicadas, con el propósito de mejorar la calidad de la educación o de investigaciones puras que buscan mejorar el entendimiento del proceso de aprendizaje (Bakhshinategh, Zaiane, ElAtia y Ipperciel, 2018). De acuerdo a lo mencionado por Candia (2019), el estudio del rendimiento académico constituye, hoy en día, un tema

destacado en la investigación educativa. Esto se debe a que, para nuestra sociedad, caracterizada por el bombardeo continuado de información desde distintos medios, el gran desafío de la educación es transformar esa gran cantidad de información en conocimiento personal válido para poder desenvolverse con eficacia en la vida. De ahí que lograr el éxito o el fracaso en los estudios es de vital importancia para el futuro profesional individual. Finalmente, Ayala *et al.* (2019), afirman que la Minería de Datos Educativa se ha posicionado como una disciplina emergente de investigación educativa que facilita el entendimiento de poblaciones estudiantiles y sus contextos. Y de manera particular, el bajo rendimiento académico se ha evidenciado como una preocupación común de las instituciones educativas.

Con base en lo anterior, es importante analizar esta información académica y para ello se hará uso de herramientas y técnicas de minería de datos educativa, lo cual permite analizar la información existente en la Universidad del Quindío que se relaciona con el rendimiento académico de los estudiantes. De esta forma, es posible apoyar los procesos de toma de decisiones orientados a proponer mecanismos que permitan implementar acciones de mejora y que de forma gradual se puedan adaptar como estrategia para favorecer los indicadores de eficiencia y efectividad. Estos indicadores son fundamentales al momento de evaluar el nivel de cumplimiento en los objetivos que en torno al rendimiento académico se han trazado tanto los estudiantes como la Universidad del Quindío.

## **MATERIAL Y MÉTODOS**

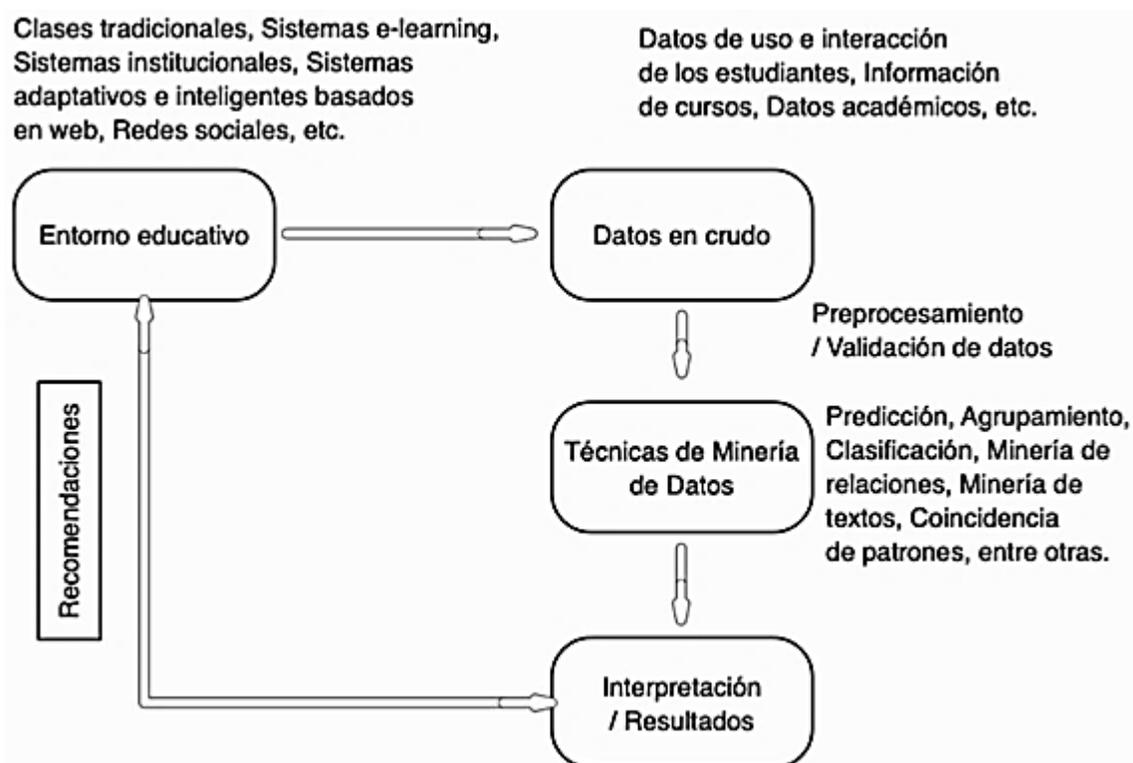
De acuerdo con Méndez (1999), un marco metodológico se refiere al conjunto de elementos de apoyos metodológicos, que se requieren para planear de la forma en que se va a proceder con el desarrollo de la investigación. Con base en los objetivos de esta investigación, se puede establecer que esta propuesta se enmarca dentro de la investigación aplicada, teniendo como fundamento lo expuesto por Valarino, Yáber y Cemborain (2010), quienes definen la investigación aplicada como una actividad que tiene por finalidad la búsqueda y consolidación del saber, así como la aplicación de los conocimientos para el acervo cultural y científico.

Teniendo como base estos referentes y los aspectos metodológicos descritos por Arroyave, Maya y Orozco (2007), se establece, de manera general, que la metodología para el desarrollo de este proyecto consta de las siguientes etapas: conocimiento, diseño y elaboración, evaluación, conclusiones y recomendaciones, documentación y presentación de la investigación.

De manera específica y considerando la existencia de métodos ampliamente aceptados que se utilizan para conducir trabajos de investigación basados en minería de datos educativos, y con el fin de aplicar la disciplina EDM basada en KDD, se utilizará el enfoque descrito en Ayala *et al.* (2019), que se ilustra en la figura 3.

**Figura 3.**

*Proceso de minería de datos educativa (EDM)*



*Nota.* Ayala *et al.* (2019).

Este enfoque se basa en un conjunto de tareas relacionadas, las cuales se describen a continuación:

- a) Extracción de datos: consiste en la elección de las variables de estudio del entorno educativo y su recuperación. Se da preferencia a datos en formato digital provenientes de sistemas informáticos diversos.
- b) Preprocesamiento: una vez recopilados los datos, se limpian o eliminan los elementos con valores redundantes, incorrectos o faltantes, se fusionan datos provenientes de múltiples fuentes y se convierten los valores a formas más apropiadas para su manejo o análisis, es decir, se validan para un manejo estandarizado y homogéneo. Esta etapa, por su complejidad, suele necesitar la aplicación de técnicas específicas para obtener datos correctamente estructurados. Es recomendable excluir variables o atributos irrelevantes, esto es, atributos que no aportan información útil para resolver el problema. También, es posible que se requiera convertir los archivos a formatos específicos del sistema empleado.
- c) Aplicación de técnicas de EDM: consiste en la selección de la técnica más adecuada dependiendo de los objetivos perseguidos (métodos de agrupamiento, clasificación, etc.). Una vez cargados los datos, pueden emplearse filtros para seleccionar atributos concretos y definir los parámetros de funcionamiento de los algoritmos para generar los modelos o descubrir patrones de los datos.
- d) Interpretación de resultados: la salida resultante de los algoritmos proporciona información que se deberá interpretar, los modelos muestran los hallazgos de patrones, relaciones o tendencias en los datos. También se pueden analizar los valores generados en la medición de la precisión de los resultados y utilizar métodos de validación. Dependiendo de la efectividad o validez de las reglas y modelos pueden requerirse más iteraciones para refinar el funcionamiento de los algoritmos (incluir otros atributos, remover atributos “ruidosos” o que no tienen peso en el modelo, aplicar filtros, etc.).
- e) Retroalimentación: el conocimiento obtenido se coloca a disposición de los actores del entorno educativo, a su vez, se pueden incluir recomendaciones; con lo

cual es posible planificar acciones para modificar o mejorar procesos o condiciones del contexto escolar.

Con respecto a la disciplina EDM en que se basa esta propuesta, se han descrito algunos métodos de análisis, que se aplican de acuerdo con las técnicas que lo sustentan. Algunas de estas técnicas son (Peña, 2014):

- a) **Predicción:** desarrollo de modelos para inferir una variable a partir de la combinación de otros datos disponibles o variables predictoras. Algunos de los métodos usados para predecir son: clasificación (cuando la variable a predecir es un valor categórico), regresión (si la variable tiene valor continuo), o estimación de densidad (cuando la variable a predecir es una función de densidad de probabilidad). Se pueden aplicar estos métodos para predecir el éxito académico de los estudiantes.
- b) **Agrupamiento:** para encontrar conjuntos de datos que se agrupen naturalmente, y separarlos del conjunto completo en una serie de categorías. Se pueden crear grupos de estudiantes basados en sus patrones de aprendizaje o estrategias cognitivas, por ejemplo.
- c) **Minería de relaciones:** se usan para descubrir relaciones entre variables y codificarlas como reglas para aplicarlas posteriormente. Algunos de sus métodos son: minería de reglas de asociación (cualquier relación entre variables), minería de patrones secuenciales (asociaciones temporales entre variables), minería de correlaciones (correlaciones lineales), minería de datos causales (relaciones causales entre variables). Usadas para identificar relaciones, de las actividades de los estudiantes con sus resultados finales, y para modelar secuencias de actividades de aprendizaje.
- d) **Descubrimiento mediante modelos:** usado para validar el modelo de un fenómeno (mediante predicción, agrupamiento o ingeniería del conocimiento). Es usado para identificar la relación entre características y comportamiento de los estudiantes.

- e) Destilado de datos: para permitir a un humano identificar o clasificar rápidamente las propiedades de un conjunto de datos. Utiliza resúmenes, visualización e interfaces interactivas para destacar la información relevante y apoyar la toma de decisiones.
- f) Detección de valores atípicos: para descubrir datos significativamente diferentes al resto del conjunto. Se puede utilizar para detectar desviaciones en las acciones o comportamientos del alumno o educador, procesos de aprendizaje irregulares y para detectar estudiantes con dificultades de aprendizaje.
- g) Análisis de redes sociales (SNA): estudia las relaciones entre los individuos. El SNA considera las relaciones sociales en términos de la teoría de red, la cual considera nodos, que representan actores individuales dentro de la red, y conexiones o enlaces, que representan relaciones entre los individuos, como amistad, relaciones cooperativas, etc.

Las técnicas que se aplicarán en este trabajo propuesto, serán definidas con base en una revisión de literatura que permita identificar cuáles son las más adecuadas para el tipo de objeto de investigación que se trata en este estudio.

## RESULTADOS

Tal como se mencionó previamente, uno de los resultados importantes de este trabajo se basa en los hallazgos a partir de la revisión de literatura. Dichos hallazgos, permiten conocer los diferentes enfoques y tratamientos que se han utilizado durante el análisis del desempeño académico estudiantil. En este sentido, existen varios trabajos orientados al análisis del fenómeno del rendimiento académico estudiantil, a través del uso de herramientas y la aplicación de minería de datos. A continuación, se relacionan algunos de ellos:

Candia (2019), presenta un trabajo con el propósito de predecir el rendimiento académico de los estudiantes. El autor analiza tanto el rendimiento académico como los factores de éxito o fracaso de los estudiantes en una institución universitaria. Como resultado,

se plantearon estrategias con el fin de mejorar las actividades que tienen que ver con los procesos de enseñanza aprendizaje.

Ayala *et al.* (2019) desarrollaron un trabajo con el propósito de encontrar factores determinantes del rendimiento académico de la población estudiantil en una universidad mexicana y específicamente en una licenciatura en Ciencias de la Computación. En este trabajo, los autores usan el método de clasificación J48, implementado en el software WEKA, para obtener árboles de decisión respecto al riesgo académico de los alumnos.

El trabajo propuesto por Rico *et al.* (2019), hace mención del uso de la minería de datos aplicada a procesos académicos, con el propósito de descubrir conocimiento y patrones a través del uso de los datos, permitiendo predecir comportamientos con base a las variables utilizadas, orientados a predecir el rendimiento académico de estudiantes universitarios mediante el algoritmo Naive Bayes.

Salal *et al.* (2019), aplicaron técnicas de minería de datos y construyeron diversos modelos de clasificación para analizar el rendimiento académico de los estudiantes. Para analizar, el conjunto de datos compuesto por 649 instancias y 33 diferentes atributos, se implementaron varios algoritmos provistos por WEKA tales como NaiveBayes, Decision Tree (J48), RandomForest, RandomTree, REPTree, JRip, OneR, SimpleLogistic y ZeroR. Los autores evaluaron y compararon los resultados de las distintas implementaciones, buscando la mejor precisión. Adicionalmente, identificaron la forma en que los atributos de los estudiantes afectan el rendimiento del desempeño de los estudiantes.

Rico y Sánchez (2018), hacen mención de la importancia de la minería de datos en el sector educativo y directamente en los procesos de enseñanza aprendizaje tendientes a fomentar propuestas y planes de prevención para los estudiantes con bajo rendimiento.

Holgado (2018) desarrolló un estudio en la Universidad Amazónica de Madre de Dios en Perú. Su desarrollo se basó en la metodología de minería de datos denominada CRISP-DM y mediante el algoritmo Random Forest permitió identificar que las variables: cantidad de asignaturas cursadas, el servicio de comedor universitario, la carrera profesional, deuda con la universidad, son las variables que más influyen en la predicción del rendimiento académico. El autor implementó tres algoritmos: Random Forest, C5.0 y CART.

Ordaz y García (2018) presentan un estudio acerca del estado del conocimiento sobre el rendimiento académico universitario como el principal indicador de problemas educativos, entre ellos el abandono. A partir de un ejercicio de análisis, en un primer momento se aborda como se ha problematizado la temática, encontrando que existe un diverso número de conceptos relacionados y variables significativas. Dichas variables se identifican a nivel del país de México y a nivel internacional.

En cuanto a la ejecución de las actividades iniciales de desarrollo del estudio de caso, se encontraron diversas fuentes de datos heterogéneas, las cuales carecen de algún mecanismo de centralización o sincronización. La información provista correspondió a los registros estudiantiles comprendidos entre los periodos 2018-1 y 2022-1. Las fuentes corresponden a la oficina de admisiones (información académica), de planeación (información socio - económica) y de estadística (información sobre resultados de pruebas de aptitudes Badyg), las cuales proporcionaron una cantidad de, 153418, 16000 y 1830 registros, respectivamente. La estructura de dicha información se ilustra en las Tablas 1, 2 y 3.

**Tabla 1.**

*Estructura de la información proporcionada por la oficina de admisiones.*

<b>Característica</b>	<b>Significado</b>
COD_UNIDAD	Código del programa académico del estudiante
COD_PENSUM	Código del pénsum académico que cursa
NUM_IDENTIFICACION	Número del documento de identificación del estudiante
NOM_LARGO	Apellidos y nombres del estudiante
COD_PERIODO	Código del periodo académico
ESTADO	Estado actual (Activo, anulado, cancelado, egresado, excluido, expulsado, graduado, inactivo, traslado)
COD_MATERIA	Código de la asignatura
NOM_MATERIA	Nombre de la asignatura

CREDITOS	Número de créditos de la asignatura
CREDITOS_MATRICULADOS	Total de créditos matriculados en el semestre
NOTA	Nota de la asignatura
PROMEDIO_SEMESTRE	Promedio del semestre
PROMEDIO_ACUMULADO	Promedio acumulado
NOM_DEPENDENCIA	Nombre de la facultad
NOM_UNIDAD	Nombre del programa académico
METODOLOGIA	Metodología presencial
MODALIDAD	Modalidad pregrado
VAL_LIQUIDADADO	Valor de matrícula liquidado
VAL_DESCTO	Valor de descuento en la matrícula
VAL_PAGADO	Valor pagado de matrícula
BAJO_ACUERDO_006	Si se encuentra bajo acuerdo 006 en ese período
BAJO_ACUERDO_005	Si se encuentra bajo acuerdo 005 en ese período

---

*Nota:* elaboración propia.

## **Tabla 2.**

*Estructura de la información proporcionada por la oficina de planeación*

<b>Característica</b>	<b>Significado</b>
ESTP_ID	Código numérico asignado por la universidad
PERIODO	Periodo correspondiente al año y semestre
UNID_NOMBRE	Sede de la universidad
COD_PROGRAMA	Código del programa académico

PROG_NOMBRE	Nombre del programa académico
JORNADA	Tipo de jornada
METODOLOGIA	Tipo de metodología
SEMESTRE	Número de semestre
PENG_PRIMERNOMBRE	Primer nombre
PENG_SEGUNDONOMBRE	Segundo nombre
PENG_PRIMERAPELLIDO	Primer apellido
PENG_SEGUNDOAPELLIDO	Segundo apellido
TIPO_DOC	Código del tipo de documento
DOCUMENTO	Número del documento
GENERO	Tipo de género
PENG_FECHANACIMIENTO	Número de día-mes-año de nacimiento
ESTADO_CIVIL	Tipo de estado civil
GRUPO_ETNICO	Tipo de grupo étnico
VÍCTIMA_CONFLICTO_ARMADO	Si es o no víctima
ES_DESPLAZADO	Si es o no desplazado
ES_DISCAPACITADO	Si es o no discapacitado
NOMBRE	Nombre de la discapacidad
ESTRATO	Número de estrato
REGIMEN_SEGURIDAD	Nombre del régimen
CATEGORIA_SISBEN	Número de la categoría
CANTIDAD_GRUPO_FAMILIAR	Cantidad de integrantes
NUCLEO_FAMILIAR	Integrantes de la familia
APORTAN_EN_LA_FAMILIA	Si aportan o no

CANTIDAD_INGRESOS	Rango de los ingresos
FUENTE_FINANCIACION	Financiación de los estudios
LABORA	Si labora o no
NOMBRE_1	Empresa donde trabaja
LIQU_TOTALLIQUIDADO	Valor de la matrícula
LIQU_TOTALDESCUENTO	Valor de descuento
LIQU_NIVELLIQUIDACION	Nivel de la liquidación
MOTIVACION	Motivo de ingreso a la universidad
ACUDIENTE	Nombre del acudiente
DIRECCIONALTERNATIVO	Dirección alterna
TELEFONOALTERNATIVO	Número de teléfono alterno

---

*Nota:* elaboración propia.

### **Tabla 3.**

*Estructura de la información proporcionada por la oficina de estadística*

<b>Característica</b>	<b>Significado</b>
CI	Coeficiente de Inteligencia
IG	Inteligencia General
RL	Razonamiento lógico
Rv	Analogías verbales
Rn	Series numéricas
Re	Matrices de figuras
Sv	Completar oraciones
Sn	Problemas numéricos

Se	Encajar figuras
Ma	Memoria auditiva
Mv	Memoria visual
De	Discriminar diferencias
RA	Rapidez
EF	Eficacia

---

*Nota:* elaboración propia.

Una vez se realizó la limpieza, preparación e integración de los datos iniciales, se obtuvo como resultado la cantidad de registros que se ilustran en la Tabla 4, para cada una de las facultades de la universidad del Quindío que hacen parte del estudio de caso.

#### **Tabla 4.**

*Cantidad de registros resultantes*

<b>Facultad</b>	<b>Cantidad de registros</b>
Ciencias Agroindustriales (2 p)	116
Ciencias Básicas y Tecnología (4 p)	295
Ciencias Económicas y Administrativas (4 p)	360
Ciencias Humanas y Bellas Artes (5 p)	101
Ciencias de la Salud (4 p)	59
Educación (6 p)	393
Ingeniería (6 p)	506

---

*Nota:* elaboración propia.

En este caso se consideró que la información de los resultados de las pruebas de aptitudes Badyg, proporcionada por la oficina de estadística, ofrecía un carácter diferenciador con todos los demás trabajos relacionados y, por lo tanto, la contribución de este estudio de caso tendría un mayor impacto en la generación de conocimiento relacionado con la temática de investigación. En consecuencia, los registros de las 3 oficinas se integraron con referencia y en torno a dicha información de la oficina de estadística.

## DISCUSIÓN Y CONCLUSIONES

Este trabajo permitió conocer diversos enfoques que se consideran relevantes para abordar la temática de estudio y análisis relacionada con el desempeño académico. En el contexto internacional se encuentra que existen varios factores generales que se tienen relevancia durante los análisis. Estos factores se agrupan en socio-demográficos, personales y previos al ingreso. De igual forma se identifican algunas variables más específicas, que tienen que ver con el rendimiento previo, asistencia a clases, rendimiento en otros cursos, prueba de admisión, motivación, escolaridad de los padres y tiempo de dedicación a los estudios, entre otros. Adicionalmente, se conoció la utilización de una gran variedad de técnicas, métodos, software y algoritmos de aprendizaje de máquina comúnmente empleadas en este tipo de estudio. Entre ellos se resaltan: método de clasificación basado en árboles de clasificación J48, Naive Bayes, RandomForest, RandomTree, C5.0, CART, REPTree, JRip, OneR, SimpleLogistic, ZeroR y software WEKA, entre otros.

Por otra parte, algunos de los resultados que se obtuvieron durante los pasos iniciales en el desarrollo del estudio de caso permiten inferir que una de las etapas decisivas y de mayor complejidad tiene que ver con la limpieza, refinamiento y estructuración de la información que servirá como insumo para los posteriores análisis. Muestra de ello radica en el gran volumen de registros con los que se contaba inicialmente, los cuales se vieron altamente disminuidos después de ejecutar dichas actividades iniciales. Lo anterior se debe a que además de la naturaleza inexacta, incompleta, ambigua y redundante de los datos, las distintas fuentes de información carecen de un repositorio central, homogéneo y sincronizado de datos.

## REFERENCIAS BIBLIOGRÁFICAS

- Aldowah, H., Al-Samarraie, H. y Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49. <https://doi.org/10.1016/j.tele.2019.01.007>.
- Arroyave, C., Maya, A. y Orozco, C. (2007). *Aplicación de QFD en el proceso de ingeniería de requisitos*. [Tesis de pregrado]. Repositorio Universidad Eafit. [https://repository.eafit.edu.co/bitstream/handle/10784/2375/Carolina\\_Arroyave\\_Carlos\\_Orozco\\_2007.pdf?sequence=3&isAllowed=y](https://repository.eafit.edu.co/bitstream/handle/10784/2375/Carolina_Arroyave_Carlos_Orozco_2007.pdf?sequence=3&isAllowed=y)
- Ayala, E. y López, R. (1-2 de julio de 2019). *Minería de datos educativa para el análisis de rendimiento académico en una carrera de computación*. [Presentación en congreso]. Congreso Nacional de Computación y Tecnología Educativa, San Luis, Argentina.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S. y 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>.
- Candia, D. (2019). *Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático*. [Tesis de maestría]. Universidad Nacional de San Antonio Abad del Cusco, Cusco, Perú.
- Holgado, L. A. (2018). *Detección de patrones de bajo rendimiento académico mediante técnicas de minería de datos de los estudiantes de la Universidad Nacional amazónica de Madre de Dios*. [Tesis de maestría]. Universidad Nacional del Altiplano, Perú.
- Kumar, M., Singh, A. J. y 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.
- Menacho, C. H. (2018). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26-33. <http://dx.doi.org/10.21704/ac.v78i1.811>.
- Méndez, C. (1999). *Metodología, diseño y desarrollo del proceso de investigación*. McGraw Hill interamericana S.A.

- Ordaz, A. y García, O. (14-15 de noviembre de 2018). El estudio del rendimiento académico en nivel universitario. Aproximaciones al estado del conocimiento. [Ponencia]. Congreso CLABES VIII, Ciudad de Panamá, Panamá.
- Peña-Ayala, A. (ed.). (2014). Educational Data Mining. *Studies in Computational Intelligence*, 524.
- Rico, A., Gaytan, N. y Sánchez, D. (2019). Construcción e implementación de un modelo para predecir el rendimiento académico de estudiantes universitarios mediante el algoritmo Naive Bayes. *Diálogos sobre educación. Temas actuales en investigación educativa*. 10(19), 1-12. <https://doi.org/10.32870/dse.v0i19.509>.
- Rico, A. y Sánchez D. (2018). Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN. *Revista iberoamericana para la investigación y el desarrollo educativo*, 8(16), 1-21.
- Salal, Y. K., Abdullaev, S. M. y Kumar, M. (2019). Educational Data Mining: Student Performance Prediction in Academic. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(4C), 54-59.
- Valarino, E., Yáber, G. y Cemborain, M. S. (2010). *Metodología de la investigación: paso a paso*. Trillas.