

MINERÍA DE DATOS PARA LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTEL UNIVERSITARIA: REVISIÓN DE LITERATURA Y EXPERIMENTACIÓN

DATA MINING FOR PREDICTING UNIVERSITY STUDENT DROPOUT: LITERATURE REVIEW AND EXPERIMENTATION

Recibido: 27/12/2024 - **Aceptado:** 18/03/2026

Marco Antonio Yandún Velasteguí

Docente de la Universidad Politécnica Estatal del Carchi
Tulcán - Ecuador

Magíster en Auditoría de Tecnologías de la Información
Universidad de Especialidades Espíritu Santo UEES

marco.yandun@upec.edu.ec

<https://orcid.org/0000-0001-5627-9838>

Carlitos Alberto Guano Cárdenas

Docente de la Universidad Politécnica Estatal del Carchi
Tulcán - Ecuador

Magíster en Gerencia de Sistemas
Universidad de las Fuerzas Armadas (ESPE)

carlos.guano@upec.edu.ec

<https://orcid.org/0000-0002-7571-2972>

Hernán Darío Cerón Carrera

Estudiante de la Universidad Politécnica Estatal del Carchi
Tulcán - Ecuador

Administrador gastronómico
Universidad UTE

hernan.ceron@upec.edu.ec
<https://orcid.org/0009-0009-1736-1949>

Alison Lizbeth Piarpuezán Imbaquingo

Estudiante de la Universidad Politécnica Estatal del Carchi
Tulcán - Ecuador

Egresada de Ingeniería en Computación
Universidad Politécnica Estatal del Carchi

alison.piarpuezan@upec.edu.ec
<https://orcid.org/0009-0008-3879-8720>

Como citar este artículo:

Yandún, M. A., Guano, C. A., Cerón, H. D., & Piarpuezán, A. L. (Julio - diciembre de 2026). Minería de datos para la predicción de la deserción estudiantil universitaria: Revisión de literatura y experimentación. *Visión Empresarial*, Vol. 16, Núm. 2, 132 – 167. <https://doi.org/10.32645/13906852.1467>



Resumen

La deserción universitaria representa un desafío crítico para las instituciones de educación superior, debido a su impacto económico, académico y social no solo en los estudiantes, sino en las instituciones. Este estudio adopta un enfoque mixto, conformado por una revisión sistemática de la literatura y una experimentación práctica mediante técnicas de minería de datos, tomando en cuenta que, en los últimos años, la minería de datos se ha convertido en una herramienta poderosa para predecir este fenómeno mediante el análisis de grandes volúmenes de datos. El objetivo de esta investigación es identificar los factores determinantes asociados al abandono estudiantil universitario y evaluar la efectividad de distintos algoritmos de aprendizaje automático para su predicción. La revisión sistemática se desarrolló bajo el método PRISMA y abarcó publicaciones de los últimos cinco años en bases de datos indexadas. Posteriormente, se ejecutó un experimento predictivo con un dataset de 34 variables y más de 4.000 registros, aplicando modelos supervisados como árboles de decisión, Random Forest y Gradient Boosting. Los resultados muestran que los factores económicos, el rendimiento académico y las condiciones institucionales, son los principales predictores de la deserción. En la experimentación, Random Forest y los árboles de gradiente obtuvieron el mejor desempeño en la predicción, mientras que los árboles de decisión ofrecieron una alta interpretabilidad útil para la toma de decisiones institucionales. Se concluye que la integración de minería de datos y modelos de aprendizaje automático permite identificar tempranamente a estudiantes en riesgo y fortalecer estrategias de retención en la educación superior.

Palabras clave: Aprendizaje automático, árboles de decisión, clasificación, predicción, deserción estudiantil universitaria.

Abstract

The university dropout rate represents a critical challenge for higher education institutions due to its economic, academic, and social impact, not only on students but also on the institutions themselves. This study adopts a mixed approach, consisting of a systematic literature review and a practical experimentation using data mining techniques, considering that in recent years, data mining has become a powerful tool for predicting this phenomenon through the analysis of large volumes of data. The objective is to identify the determining factors associated with university student dropout and to evaluate the effectiveness of different machine learning algorithms for its prediction. The systematic review was conducted following the PRISMA method and included publications from the last five years in indexed databases. Subsequently, a predictive experiment was carried out using a dataset with 34 variables and more than 4,000 records, applying supervised models such as decision trees, Random Forest, and Gradient Boosting. The results show that economic factors, academic performance, and institutional conditions are the main predictors of dropout. In the experimental phase, Random Forest and Gradient Boosting yielded the best predictive performance, while decision trees offered high interpretability useful for institutional decision-making. The study concludes that the integration of data mining and machine learning models enables early identification of at-risk students and strengthens retention strategies in higher education.

Keywords: Machine learning, decision trees, classification, prediction, university student dropout

Código JEL: O32 Gestión de la Innovación Tecnológica y la Investigación y Desarrollo

Introducción

La deserción universitaria continúa siendo uno de los desafíos más complejos para las instituciones de educación superior, no solo por las pérdidas económicas y académicas que genera, sino también por el impacto humano que deja en la vida de miles de estudiantes. Detrás de cada caso existe una historia marcada por dificultades económicas, problemas familiares, escasa preparación académica o falta de acompañamiento institucional. En América Latina, estas realidades se han intensificado en los últimos años, lo que evidencia la necesidad de comprender este fenómeno con mayor profundidad (Gutiérrez et al., 2024).

En Ecuador, este fenómeno es crítico. Antecedentes recientes indican tasas de abandono de hasta el 58.95% en sectores de bajos ingresos y superiores al 52% durante los primeros ciclos académicos. Estas cifras evidencian que las estrategias actuales de retención resultan insuficientes para frenar la problemática. En este contexto, la minería de datos y las técnicas de aprendizaje automático han surgido como enfoques prometedores para analizar patrones complejos y predecir la deserción con alta precisión.

Diversos estudios coinciden en que abandonar la universidad no es un suceso aislado ni repentino. Más bien, es el resultado de una combinación de factores que interactúan entre sí: el desempeño académico, el bienestar emocional, la situación económica y el entorno familiar, entre otros. Sin embargo, comprender esta interacción no siempre es sencillo y los métodos tradicionales de análisis suelen quedarse cortos para detectar patrones complejos o comportamientos que predicen el riesgo de abandono (Galve, 2024).

La necesidad de esta investigación radica en la urgencia de adoptar un enfoque preventivo mediante el uso de tecnología. Aunque la literatura ha identificado múltiples causas del abandono, existe un vacío en la aplicación de herramientas que permitan anticiparlo. Por ello, este estudio justifica la implementación de técnicas de minería de datos y aprendizaje automático, capaces de procesar grandes volúmenes de información para detectar patrones de riesgo que los métodos tradicionales no logran visualizar.

En este escenario, la minería de datos ha ganado un rol fundamental. Gracias a su capacidad para procesar grandes volúmenes de información, permite descubrir relaciones que no son visibles a simple vista y anticipar situaciones que podrían afectar la continuidad educativa de los estudiantes. Esto resulta especialmente valioso para países como Ecuador, donde factores como la necesidad de trabajar, la falta de recursos y el bajo rendimiento continúan influyendo de manera significativa en la permanencia universitaria (Muñoz, 2024).

A pesar de estos avances, aún existen vacíos en la literatura, especialmente en estudios que integren de manera conjunta los aspectos académicos, socioeconómicos y emocionales, y que además validen los modelos predictivos con datos reales. Esta ausencia evidencia la necesidad de investigaciones más completas, que no solo analicen lo que se ha dicho en los últimos años, sino que también lleven esos conocimientos a la práctica para generar herramientas útiles y aplicables (Barahona-Aguisaca et al., 2024).

El objetivo principal es identificar los factores determinantes de la deserción y evaluar la efectividad de diversos algoritmos predictivos. Se plantea como hipótesis que la integración de modelos de clasificación, comparando la precisión de Random Forest con la interpretabilidad de los

Árboles de Decisión, permite identificar tempranamente a los estudiantes en riesgo, de modo que facilita la creación de estrategias de retención oportunas y fundamentadas en datos.

Marco teórico

Definición y Aplicaciones

- o. **Minería de datos.** La minería de datos es un proceso que permite descubrir patrones y conocimiento oculto en grandes volúmenes de datos. Sus aplicaciones en la educación incluyen el análisis del rendimiento académico, la predicción de abandono y la identificación de factores de riesgo.
- p. **Técnicas Comunes en Minería de Datos.** La minería de datos, como subcampo de la inteligencia artificial, proporciona herramientas analíticas robustas para extraer patrones significativos de grandes volúmenes de datos educativos (Luna-Rizo et al., 2024), lo cual facilita la predicción de la deserción universitaria (Noriega et al., 2024). En el contexto de la retención estudiantil, estas técnicas permiten modelar variables como el rendimiento académico, datos demográficos y patrones de interacción en sistemas de gestión del aprendizaje (LMS), con el fin de identificar estudiantes en riesgo de abandono de manera temprana (Osorio, 2024). Estudios como el de Guadalupe Mori (2025), destaca la integración de métodos supervisados (clasificación y regresión) y no supervisados (clustering) para mejorar la precisión predictiva. Con ello, supera enfoques tradicionales estadísticos al manejar la complejidad y no linealidad inherente a los datos educativos. A continuación, se detalla las técnicas más comunes aplicadas en este tema.
- q. **Clasificación: Árboles de decisión, SVM, redes neuronales.** La clasificación supervisada se utiliza para predecir la deserción, tomando en cuenta el resultado final del estudiante en relación con su condición académica (abandono, retención y graduación) (Mazurett González, 2025). Esto se logra utilizando algoritmos que aprenden de datos etiquetados para asignar categorías a nuevas instancias. Los árboles de decisión, por su interpretabilidad y capacidad para manejar variables categóricas y continuas, son particularmente útiles en entornos educativos (Párraga, 2024). Estos modelos dividen el espacio de características en subregiones basadas en reglas de umbral, generando un diagrama de flujo que facilita la identificación de factores clave de acuerdo con el dataset utilizado.
- r. **Regresión: Modelos de regresión logística y lineal.** La regresión, como técnica supervisada, modela la relación entre variables predictoras y una respuesta continua o categórica, siendo esencial para estimar probabilidades de deserción. En el estudio de Rabelo y Zárate (2024) indican que la regresión logística —una extensión de la regresión lineal para outcomes binarios estima la logit de la probabilidad de abandono mediante la función sigmoide, facilitando interpretaciones probabilísticas. En el análisis de retención universitaria, se ha clasificado correctamente el 89% de los casos, identificando el 98.1% de desertores al integrar ensambles con redes neuronales y árboles (Ujkani, Minkovska y Stoyanova, 2022).

Clustering: K-means y DBSCAN

El *clustering* no supervisado agrupa estudiantes sin etiquetas previas, revelando perfiles de riesgo latentes basados en similitudes en datos de comportamiento y rendimiento. K-means, un algoritmo particional, minimiza la varianza intra-cluster asignando puntos a k centroides iterativamente, siendo efectivo para segmentar estudiantes en grupos de bajo, medio y alto riesgo de deserción. Aplicado a datos de admisiones, generó clústeres con siluetas de 0.575, de modo que identificó perfiles de desertores con precisión del 80%. Su eficiencia escalable lo hace ideal para grandes cohortes universitarias (Peet Moraga, 2025).

Algoritmos y enfoques en la predicción de deserción

La predicción de la deserción universitaria mediante minería de datos y aprendizaje automático ha evolucionado rápidamente, incorporando algoritmos supervisados que analizan patrones en datos académicos, sociodemográficos y de comportamiento para identificar estudiantes en riesgo con alta precisión (Passeggia et al., 2025). Entre los enfoques más prometedores se encuentran los modelos de ensemble y boosting, como XGBoost y CatBoost, que destacan por su robustez en conjuntos de datos desbalanceados y su capacidad para manejar variables categóricas sin preprocesamiento extenso. La optimización multiobjetivo mediante el algoritmo genético NSGA-II, combinado con CatBoost y técnicas de balanceo como ADASYN, mejoró la predicción de deserción en logs de Moodle Rebelo et al., (2025). Gracias a esto, fue posible lograr un F1-score de 0.86 y AUC de 0.95 en la semana 7 de cursos, superando a modelos base como Random Forest y XGBoost en generalización a cursos no vistos.

Nomenclatura y definiciones de términos. Con el propósito de mantener consistencia conceptual y terminológica a lo largo del estudio, se establecen las siguientes definiciones y abreviaturas utilizadas en el análisis de minería de datos y aprendizaje automático aplicado a la predicción de la deserción universitaria.

Deserción universitaria (Dropout): se refiere al abandono definitivo o temporal de los estudios por parte de un estudiante antes de completar su programa académico. En este trabajo los términos deserción, abandono estudiantil y dropout se utilizan como equivalentes.

Dataset: conjunto estructurado de datos utilizado para el entrenamiento y evaluación de los modelos de aprendizaje automático.

Training set: subconjunto de datos utilizado para entrenar los algoritmos de aprendizaje automático.

Test set: subconjunto de datos utilizado para evaluar el rendimiento del modelo entrenado.

Decision Tree (DT): algoritmo de clasificación supervisada que estructura las decisiones en forma de árbol basado en reglas de división de variables.

Random Forest (RF): método de aprendizaje ensemble basado en múltiples árboles de decisión que mejora la precisión mediante la combinación de predicciones.

Gradient Boosting (GB): técnica de aprendizaje ensemble que construye modelos de forma secuencial para reducir el error del modelo anterior.

Variable objetivo (Target): variable dependiente utilizada para clasificar a los estudiantes en tres categorías: Dropout, Graduate y Enrolled.

En el presente estudio se utilizará una notación matricial para representar el conjunto de datos, donde $X \in \mathbb{R}^{n \times m}$ representa la matriz de características, n corresponde al número de observaciones y m al número de variables analizadas.

Materiales y métodos

Diseño de la investigación

El presente estudio adopta un enfoque de investigación mixto de tipo secuencial exploratorio, que integra una fase cualitativa basada en una revisión sistemática de la literatura y una fase cuantitativa de experimentación con modelos de aprendizaje automático. En la primera fase se realizó una revisión sistemática siguiendo el protocolo PRISMA, con el objetivo de identificar los principales factores asociados a la deserción estudiantil universitaria y las técnicas de minería de datos más utilizadas para su predicción.

Los resultados obtenidos en esta fase permitieron construir una base conceptual para el estudio, identificando las variables más relevantes reportadas en la literatura, tales como factores académicos, socioeconómicos, institucionales y personales. Esta información sirvió como fundamento teórico para orientar la fase cuantitativa del estudio.

En la segunda fase se desarrolló una experimentación con modelos de aprendizaje automático utilizando un conjunto de datos de estudiantes universitarios. La selección de variables, así como la elección de algoritmos de clasificación (árboles de decisión y Random Forest), se fundamentó en los hallazgos de la revisión sistemática, lo que permitió validar empíricamente los factores identificados en la literatura.

De esta manera, la integración de ambas fases permitió no solo comprender el fenómeno de la deserción desde una perspectiva teórica, sino también evaluar su comportamiento mediante modelos predictivos aplicados a datos reales, fortaleciendo la validez del estudio.

Revisión sistemática de literatura

Para la identificación de los factores de deserción, se aplicó el método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

Preguntas de investigación

La revisión se orientó a responder las siguientes interrogantes planteadas para guiar el estudio:

P1: ¿Cuáles son los factores que influyen en la deserción de estudiantes universitarios?

P2: ¿Qué consecuencias tiene la deserción en la vida de los estudiantes?

P3: ¿Cómo afecta la deserción a las instituciones públicas de educación superior?

P4: ¿Qué técnicas de minería de datos se puede aplicar para la predicción de deserción estudiantil universitaria?

Estrategia de búsqueda

Las búsquedas se realizaron en las bases de datos Scopus, Web of Science, IEEE Xplore y Google Scholar, considerando publicaciones entre 2019 y 2024, con el objetivo de identificar estudios relacionados con la aplicación de técnicas de minería de datos y aprendizaje automático para la predicción de la deserción estudiantil en educación superior. Con ello, busca identificar patrones, tendencias y relaciones en conjuntos de datos estructurados y no estructurados. En el ámbito universitario, para su predicción debe aplicarse algoritmos de aprendizaje automático y técnicas de análisis de datos, apoyándose en patrones como el comportamiento y rendimiento académico que permitan predecir la deserción, identificación de factores de riesgo, segmentación de estudiantes, trayectoria académica.

Ecuación de búsqueda

Con el fin de garantizar la replicabilidad del proceso de revisión sistemática, se definió una ecuación de búsqueda estructurada utilizando operadores booleanos. La estrategia combinó términos relacionados con la deserción universitaria, minería de datos y aprendizaje automático. La ecuación aplicada en las bases de datos fue la siguiente:

“student dropout” OR “university dropout” OR “academic dropout” OR “student attrition”) AND (“data mining” OR “machine learning” OR “educational data mining”) AND (“prediction” OR “classification model” OR “predictive model”) AND (“higher education” OR “university students”)

Las búsquedas se realizaron en las bases de datos Scopus, Web of Science, IEEE Xplore y Google Scholar, considerando publicaciones entre 2019 y 2024, con el objetivo de identificar estudios relacionados con la aplicación de técnicas de minería de datos y aprendizaje automático para la predicción de la deserción estudiantil en educación superior.

Criterios de inclusión y exclusión

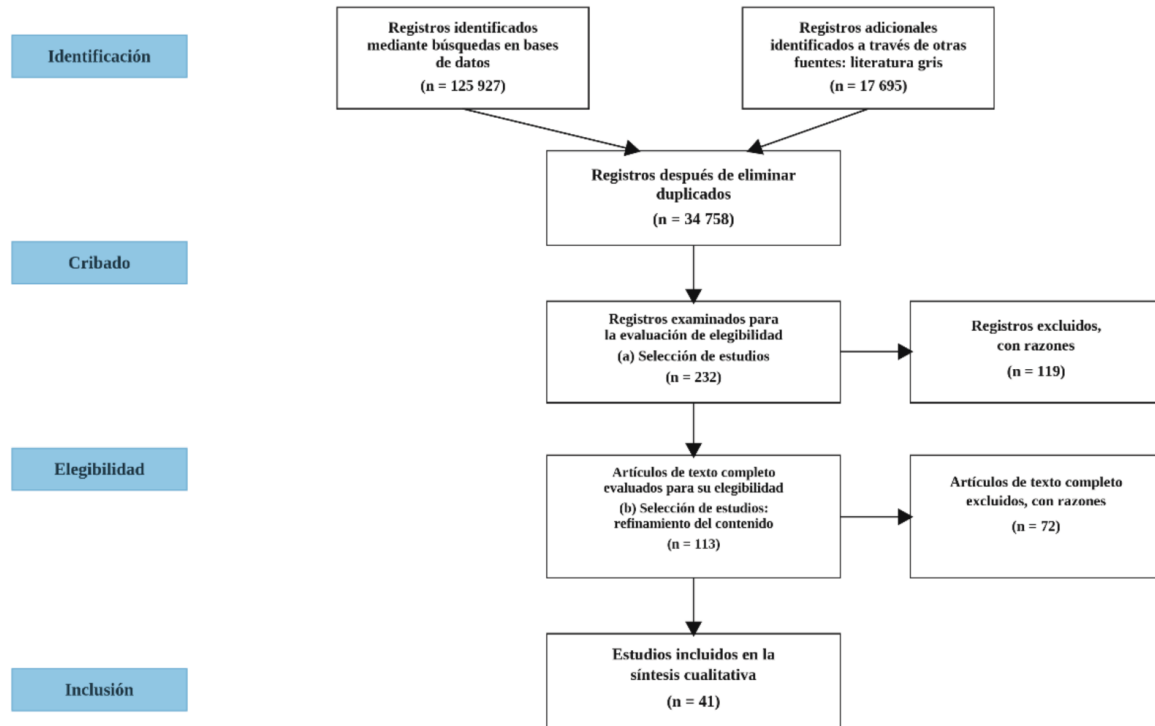
Inclusión: Artículos publicados en los últimos cinco años, estudios empíricos, revisiones sistemáticas y experimentos con modelos predictivos.

Exclusión: Artículos no relacionados con predicción de deserción, publicaciones no revisadas por pares, y estudios con datos insuficientes y población de estudio diferente.

Proceso de Selección

La selección de artículos se realizó siguiendo el método PRISMA, que incluye la identificación, cribado y selección final de estudios relevantes.

Figura 1
Aplicación de metodología PRISMA



Población y objeto de estudio

La experimentación no requirió la recolección de datos primarios en campo, sino que utilizó una fuente secundaria validada. El objeto de estudio fue el conjunto de datos “*Predict Students’ Dropout and Academic Success*”, obtenido del repositorio UCI Machine Learning Repository.

Características de la muestra

El dataset consta de 4,424 registros (estudiantes) y 34 variables.

Variables

Incluye datos demográficos (estado civil, nacionalidad, edad), socioeconómicos (ocupación de los padres, becas), y académicos (calificaciones previas, unidades curriculares aprobadas).

Preprocesamiento

Los datos se encontraban normalizados y las variables categóricas, como el estado civil, codificadas numéricamente para facilitar su procesamiento algorítmico.

Procedimiento y métodos de análisis

Para la construcción del modelo predictivo se aplicaron técnicas de aprendizaje supervisado, específicamente algoritmos de clasificación. El procedimiento metodológico siguió los siguientes pasos:

- a. **Carga y limpieza:** Importación del archivo *DATA_EXPERIMENTO_DESECCION.csv* y verificación de tipos de variables.
- b. **Partición de datos:** Se dividió el conjunto de datos en dos subconjuntos: 80% para el entrenamiento del modelo (*training set*) y 20% para la prueba (*test set*), utilizando una semilla aleatoria (*random_state=42*) para garantizar la reproducibilidad de los resultados.
- c. **Entrenamiento del modelo:** Se configuró un clasificador de Árbol de Decisión (*DecisionTreeClassifier*), limitando la profundidad del árbol (*max_depth=4*) para evitar el sobreajuste y mejorar la interpretabilidad del modelo. Adicionalmente, se exploró el uso de *Random Forest* para comparar métricas de precisión.
- d. **Evaluación:** Se evaluó el desempeño del modelo en su capacidad para clasificar a los estudiantes en tres categorías objetivo: *Dropout* (Deserción), *Graduate* (Graduado) y *Enrolled* (Matriculado).

Resultados y discusión

Factores que influyen en la deserción de estudiantes universitarios

Factores económicos. Falta de recursos financieros: La incapacidad para costear matrículas, transporte, materiales y otros gastos asociados a la universidad es una causa recurrente de la deserción estudiantil. Estudios destacan que la pérdida de empleo o la ausencia de apoyo económico familiar agravan esta situación, como muestra Vélez et al. (2025) en su estudio, indican que uno de los factores más significativos que contribuye a este problema, está vinculado al ámbito socioeconómico el cual se conecta con otras dificultades como el desempleo. Estos elementos se transforman en obstáculos que dificultan la continuidad en los procesos de educación superior, ya que los estudiantes enfrentan prioridades diarias que deben atender de manera inmediata. El estudio realizado por Núñez (2024) evidencia que los estudiantes que pertenecen a grupos con ingresos más bajos presentan tasas de deserción considerablemente altas, alcanzando un 58.95 %.

Necesidad de trabajar: Muchos estudiantes abandonan sus estudios para incorporarse al mercado laboral y generar ingresos. Según Vélez et al. (2025), la necesidad de contribuir al sustento familiar obliga a los estudiantes a priorizar el empleo sobre la educación, especialmente en hogares de bajos ingresos donde los jóvenes son vistos como proveedores económicos. En Ecuador, la precariedad laboral y los bajos salarios dificultan la conciliación entre trabajo y estudio, lo que lleva a una interrupción permanente de la trayectoria académica.

Factores académicos. Bajo rendimiento académico: Las calificaciones bajas, la repetición de asignaturas o dificultades para adaptarse a las exigencias universitarias son predictores clave de deserción. En la investigación realizada por Buenaño et al. (2024) se estima que el 52,98% de los estudiantes abandonan la carrera en el primer ciclo debido a la pérdida de materias, lo que confirma que las calificaciones bajas y las dificultades para adaptarse a las demandas universitarias son indicadores clave de la deserción. Por otra parte, en Ecuador la investigación de Orellana et al. (2024) indica que el bajo rendimiento es una de las principales causas de abandono de los estudios y concuerda con el estudio realizado por Katt et al. (2024) donde los resultados académicos

mostraron un elevado porcentaje de reprobación, ya que más del 50% de las asignaturas no fueron aprobadas, siendo esta la principal causa de la deserción.

Falta de preparación previa: Una formación deficiente en niveles educativos anteriores dificulta el éxito en la universidad. Núñez (2024) sugiere que la deserción durante los primeros semestres de la carrera se debe a aspectos vinculados con la preparación académica recibida antes de ingresar a la universidad. La brecha significativa entre la educación secundaria y la superior es un obstáculo especialmente en los estudiantes que consideran que su formación secundaria no les proporcionó una base sólida en los contenidos, lo que dificultó su adaptación académica y provocó incertidumbre desde el inicio sobre su capacidad para continuar con la carrera (Fonseca et al. 2022).

Expectativas no cumplidas: Cuando la carrera no satisface las expectativas vocacionales o profesionales, los estudiantes tienden a abandonar sus estudios. Fonseca et al. (2022) mencionan que los estudiantes que tuvieron problemas para adaptarse al entorno académico universitario crearon una sensación de desaprendizaje, confusión, frustración y sensación de rezago, lo que resultó en desmotivación y una mayor percepción de obstáculos y presiones. Esta situación provocó que sus expectativas iniciales sobre la universidad no se cumplieran y terminaran por abandonar la carrera.

Factores personales y psicológicos. Falta de motivación: La ausencia de un proyecto de vida claro o desinterés por los estudios influye en la decisión de desertar. Los estudiantes que no tienen claro su proyecto de vida pueden llegar a sentir esa confusión sobre su futuro, es una de las razones principales por las que deciden dejar la universidad (Vidal Chica 2023).

Problemas emocionales: Los problemas emocionales y los trastornos psicológicos, como la ansiedad, la depresión y el estrés, pueden afectar de manera significativa la capacidad de los estudiantes para concentrarse y mantenerse motivados en sus actividades académicas, Jarrin y Moreta (2024). La presión académica junto con la falta de apoyo psicológico incrementa estos problemas emocionales, lo que disminuye la autoestima de los estudiantes y contribuye a que muchos decidan abandonar la universidad. Por lo tanto, estas dificultades no solo afectan el rendimiento académico, sino que también influyen en la continuidad educativa (Orellana et al., 2024).

Orientación vocacional deficiente: Los estudiantes consideran la orientación profesional como un factor significativo para continuar con los estudios universitarios, Gracia y Vega (2024). En este sentido, Quincho (2024) enfatiza que una orientación adecuada fomenta una elección de carrera consciente, alineando los intereses y capacidades de los estudiantes con sus metas profesionales, lo que fortalece su motivación y compromiso académico. Además, Duque et al. (2024) sugieren que la intervención de un profesional en orientación vocacional antes del ingreso a la educación superior es fundamental para reducir el riesgo de deserción, al evitar decisiones erróneas en la selección de carrera que puedan derivar en abandono.

Factores familiares y sociales. Falta de apoyo familiar: La ausencia de soporte emocional o económico incrementa el riesgo de deserción. Acosta et al. (2024) mencionan que la falta de motivación por parte de los padres es un factor crucial, ya que afecta la autoestima de sus hijos, dificultando que estos logren completar sus estudios universitarios. En el estudio se determina que este factor afecta al 62% de los estudiantes. En este mismo contexto, Conde et al. (2023) señalan que el alto índice

de abandono escolar está influenciado por las condiciones sociofamiliares que afectan las actitudes y conductas de los estudiantes en el aula.

Integración social: Las dificultades para adaptarse al ambiente universitario, como los conflictos sociales y las exigencias académicas, son factores determinantes en la deserción estudiantil. Según Escalante et al. (2023), las demandas académicas y las estrategias de aprendizaje desempeñan un papel clave, ya que los estudiantes deben ajustarse a un entorno más exigente, lo que lleva a muchos a abandonar sus estudios al no poder manejar la presión de la formación profesional. Por su parte, Orellana et al. (2024) complementan esta idea al señalar que la deserción es más frecuente en los primeros semestres y se asocia a la difícil transición de la educación media a la superior, donde los estudiantes enfrentan mayores responsabilidades y desafíos académicos.

Factores institucionales. Calidad de la educación: Los planes de estudio desactualizados, las metodologías poco efectivas o la falta de apoyo docente impactan negativamente en la motivación de los estudiantes. Según Núñez (2024), en Ecuador, los factores instruccionales y académicos desempeñan un papel determinante en este problema y a pesar de los esfuerzos por mejorar la calidad educativa, persisten desafíos significativos en la adaptación académica y en la calidad de la enseñanza, los cuales impactan directamente la permanencia estudiantil.

Falta de acompañamiento: La ausencia de tutorías, asesorías o programas de retención dificulta la continuidad de los estudiantes en sus estudios de educación superior. Villegas y Núñez (2024) explican que la falta de programas de acompañamiento, como tutorías y asesorías personalizadas, es un factor crítico que contribuye a la deserción universitaria en América Latina. Los autores argumentan que las instituciones que no implementan estrategias de retención efectivas tienen un aumento de abandono, especialmente entre estudiantes de primer ingreso.

Gestión institucional: Factores como horarios flexibles, infraestructura adecuada, programas de orientación profesional y apoyo socioeconómico, son esenciales para reducir la deserción. Por ende, la deficiencia en estos procesos conduce a los estudiantes al desapego institucional, Delgado (2024). Según Quincho (2024), los aspectos institucionales, como la calidad de los servicios de apoyo y las políticas educativas adecuadas, son fundamentales para garantizar el éxito académico de los estudiantes y mejorar la retención estudiantil.

Los factores económicos tienen una alta incidencia en la deserción. Acosta-Román et al. (2024) muestran que, en Perú, el 42 % de los estudiantes están en alto riesgo de deserción por motivos económicos, mientras que, en Ecuador, Loaiza et al. (2023) identifican las dificultades económicas como el factor más significativo, seguido por problemas académicos y personales. En la Tabla 1 se categorizan las variables junto con la valoración cuantificada, según la escala de Likert, basada en la frecuencia e impacto reportados en los estudios mencionados previamente.

Tabla 1

Factores que influyen en la deserción universitaria y sus referencias

Categoría	Factor	Valoración (1-5)	Referencias
Económicos	Falta de recursos financieros	5	Vélez et al., (2025), Núñez (2024)
	Necesidad de trabajar	4	Vélez et al., (2025)
Académicos	Bajo rendimiento académico	4	Castro et al., (2021); Orellana et al., (2024); Katt et al., (2024)
	Falta de preparación previa	3	Núñez, (2024); Fonseca Et al., (2022)
	Expectativas no cumplidas	3	Fonseca et al., (2022)
Personales y psicológicos	Falta de motivación	3	Vidal Chica, (2023)
	Problemas emocionales	3	Jaramillo y Moreta (2024); Orellana et al., (2024)
	Orientación vocacional deficiente	3	De Gracia y Vega (2024); Quincho (2024); Duque et al., (2024)
Familiares y sociales	Falta de apoyo familiar	4	Acosta et al., (2024); Conde et al., (2023)
	integración social	3	Escalante et al., (2023); Orellana et al., (2024)
Institucionales	Calidad de la Educación	3	Núñez (2024)
	Falta de acompañamiento	4	Villegas y Núñez (2024)
	Gestión institucional	3	Quincho (2024); Delgado (2024)

Consecuencias de la deserción en la vida de los estudiantes

Impacto en la empleabilidad. La deserción universitaria en Ecuador incide directamente en las oportunidades laborales de los estudiantes, al limitar su acceso a empleos formales y bien remunerados. En un entorno laboral cada vez más competitivo, la falta de un título universitario coloca a los desertores en situaciones de vulnerabilidad laboral. Rojeab et al. (2024) afirman que, en ciertas carreras técnicas existe una saturación del mercado, lo cual genera desempleo y desincentiva la continuidad de los estudios superiores. Esta dinámica se convierte en un círculo vicioso: abandonar

la universidad en busca de ingresos inmediatos solo lleva a empleos informales o precarios. Además, Vélez et al. (2025) explican que los factores socioeconómicos son determinantes en la deserción, ya que muchos estudiantes deben abandonar sus estudios para apoyar económicamente a sus familias. Esta necesidad de generar ingresos a corto plazo sacrifica el acceso a una formación académica que, a mediano y largo plazo, garantizaría una mejor empleabilidad.

Limitación en la movilidad social. Abandonar los estudios universitarios limita severamente las posibilidades de los jóvenes ecuatorianos para mejorar su situación socioeconómica. Según el Instituto Nacional de Estadística y Censos (INEC, 2024), quienes no culminan la educación superior tienen mayores probabilidades de quedar atrapados en condiciones de subempleo o pobreza. Vélez et al. (2025) también coinciden en que la deserción universitaria actúa como un obstáculo para la movilidad social ascendente, ya que impide la obtención de credenciales académicas que son claves para acceder a mejores oportunidades.

Impacto psicológico y emocional. La dimensión emocional y psicológica es un factor muchas veces invisibilizado en el análisis de la deserción. En este contexto, Moreta-Herrera et al. (2021) evidencian que una parte significativa de la población universitaria en Ecuador presenta síntomas relevantes de ansiedad y depresión. Estas condiciones de malestar psicológico no solo afectan el bienestar personal del estudiante, sino que pueden agravar su sensación de fracaso y convertirse en detonantes decisivos para el abandono de la carrera. Estos efectos psicológicos no solo afectan su bienestar inmediato, sino que también disminuyen su capacidad para retomar la educación o insertarse de manera exitosa en el ámbito laboral. Asimismo, Muñoz y Cedeño (2024) señalan que la soledad no deseada, un fenómeno creciente en el entorno universitario ecuatoriano, contribuye a la desconexión emocional con el entorno académico, afectando la motivación y la permanencia estudiantil.

Efecto en el desarrollo personal y profesional. La interrupción de los estudios universitarios impide el desarrollo integral de los jóvenes, tanto en el plano personal como en el profesional. Villegas y Núñez (2024) sostienen que la ausencia de programas institucionales de acompañamiento estudiantil debilita la permanencia y limita la formación de competencias clave para el mundo del trabajo. La falta de estrategias de retención impide que los estudiantes completen su proceso de formación, afectando su autoconfianza y su preparación para afrontar los retos profesionales. Por otro lado, Ávila et al. (2024) advierten que las falencias en la gestión institucional —como la escasez de tutorías, asesorías académicas o servicios de orientación vocacional— también inciden en la decisión de abandonar los estudios. Estas carencias frenan el potencial de desarrollo de los estudiantes y reducen su capacidad de proyección a futuro, en la Tabla 2 se muestra los resultados de estudios en donde se realiza una valoración de acuerdo con el factor y la justificación literaria encontrada.

Tabla 2

Categorías, factores y justificaciones de la deserción universitaria (2025)

Categoría	Valoración (1-5)	Factor	Justificación
Empleabilidad	5	Saturación del mercado laboral	El exceso de profesionales en ciertas carreras genera desempleo, lo que desincentiva la continuidad de los estudios (Granda et al., 2024).
	4	Factores socioeconómicos	La necesidad de trabajar para sostener a la familia limita la posibilidad de mantenerse en la universidad (Vélez et al., 2025).
Movilidad social	4	Limitación en la movilidad social	El abandono universitario perpetúa la pobreza al restringir las oportunidades de mejora social (Vélez et al., 2025).
	3	Desigualdad en zonas rurales	Jóvenes en zonas rurales tienen menor acceso a educación superior, reduciendo sus posibilidades de ascenso social (INEC, 2024).
Impacto psicológico y emocional	4	Problemas emocionales	La ansiedad y la depresión son frecuentes entre los universitarios Moreta-Herrera et al. (2021).
	3	Soledad no deseada	La falta de integración social universitaria incrementa el aislamiento y disminuye la motivación (Muñoz y Cedeño, 2024).
Efecto en el desarrollo personal	4	Falta de acompañamiento	Programas de retención son clave para reducir la deserción y fomentar la permanencia (Rafael Villegas y Núñez Lira, 2024).
	3	Gestión institucional	Deficiencias en la administración académica, como falta de tutorías o servicios, influyen en el abandono (Ávila et al., 2024).

Consecuencias de la deserción en las instituciones públicas de educación superior

Tabla 3

Afectaciones de la deserción universitaria y referencias

Autores	Afectación	Artículo
Amaya et al., 2020; Salinas, 2023	Compromete la sostenibilidad financiera de las universidades Pérdida de recursos públicos invertidos en estudiantes	Big data, una estrategia para evitar la deserción escolar en las IES. Deserción en la Educación Superior en Ecuador, causas y consecuencias. Crisis económica post pandemia: cómo afecta en el acceso a la educación superior en el Paraguay.
Herreño-Munera et al., 2024; Cajigal-Molina, 2022	Deficiencias en calidad educativa, afecta reputación y atracción de alumnos	La deserción estudiantil en la educación superior: tendencias y oportunidades en la era post pandemia. Resiliencia y deserción escolar. Un estudio para plantear estrategias desde la tutoría en la educación superior.
Castro Guerra, 2023; Salcedo-Escarria, 2020	Rediseño de procesos y estrategias de retención con más recursos	Factores que influyen en la deserción estudiantil en Instituciones de Educación Superior en Latinoamérica durante los años 2017 a 2022: una revisión de literatura. Deserción universitaria en Colombia.
Terrazas et al., 2022; López Acosta, 2023; Ministerio de Educación Nacional, 2022	Frustra movilidad social, barreras por pandemia	El estrés académico y afectaciones emocionales en estudiantes de nivel superior. Estrategias para promover la permanencia en educación superior.

La deserción estudiantil en las instituciones públicas de educación superior genera efectos significativos en múltiples dimensiones. Económicamente, implica una pérdida de recursos públicos invertidos en estudiantes que no culminan sus estudios, incluyendo subsidios estatales y costos operativos, lo que compromete la sostenibilidad financiera de las universidades (Amaya et al., 2020; Granda, 2024; Salinas, 2023). Académicamente, altas tasas de deserción reflejan deficiencias en la calidad educativa, como la falta de preparación previa de los estudiantes o insuficientes programas de acompañamiento, lo que afecta la reputación institucional y su capacidad para atraer nuevos alumnos (Cajigal-Molina, 2022; Herreño-Munera et al., 2024).

Desde el ámbito organizativo, en un estudio realizado en Colombia se determinó que la deserción obliga a las instituciones a rediseñar procesos administrativos y académicos, implementando

estrategias de retención que demandan recursos adicionales, como tutorías y servicios psicológicos (Castro Guerra, 2023; Salcedo-Escarria, 2020). Socialmente, el abandono estudiantil frustra las expectativas de movilidad social de los estudiantes y limita la formación de profesionales necesarios para el desarrollo económico, un desafío crítico para las universidades públicas como pilares de equidad (López Acosta, 2023; Terrazas et al., 2022). Finalmente, un estudio realizado en Colombia demostró que la pandemia por COVID-19 agudizó estos problemas al incrementar las barreras tecnológicas y económicas, forzando a las instituciones a innovar en inclusión digital y bienestar estudiantil (Ministerio de Educación Nacional, 2022).

La deserción estudiantil tiene un impacto significativo en las universidades públicas, ya que provoca pérdidas económicas debido al desperdicio de recursos, afecta su reputación académica, obliga a modificar procesos internos para mantener la permanencia estudiantil y limita las oportunidades de movilidad social de los jóvenes. Por esta razón, es fundamental que estas instituciones adopten estrategias prácticas y eficaces, como programas de tutoría, asistencia psicológica y acceso a tecnologías digitales, con el fin de apoyar a los estudiantes y asegurar que concluyan sus estudios satisfactoriamente.

Técnicas de minería de datos para la predicción de deserción estudiantil universitaria

La revisión sistemática de la literatura revela que existen diversas técnicas de minería de datos efectivas para predecir la deserción estudiantil universitaria, permitiendo identificar patrones y factores de riesgo de manera temprana. Según Cedillo et al. (2024), la Minería de Datos Educativa (MDE) emplea métodos como estadísticos y algebraicos, redes bayesianas, algoritmos evolutivos, tablas de contingencia, árboles de decisión, redes neuronales artificiales, y enfoques estocásticos y borrosos para analizar datos educativos y predecir resultados como la deserción, considerando variables como rendimiento académico y comportamiento estudiantil. Por su parte, Parra et al. (2023) destacan el uso de árboles de decisión, métodos probabilísticos, regresión logística, redes neuronales, k-means y k-vecinos, señalando que los árboles de decisión son los más utilizados por su eficacia en la predicción.

Estos métodos mencionados anteriormente se complementan con algoritmos como: DecisionTable, JRip, OneR, PART, DecisionStump, HoeffdingTree, J48, LMT, Random- Forest, RandomTree y REPTree, aplicados en el estudio de Cornejo et al. (2023). En este se concluye que RandomTree fue el más preciso para predecir la deserción, identificando factores clave como el semestre cursado, exámenes extraordinarios, región de procedencia, edad y calificación de admisión. Finalmente, Melo (2024) utilizó algoritmos de aprendizaje supervisado como regresión logística, árboles de decisión, Random Forest, K-Nearest Neighbors (KNN) y Support Vector Machine (SVM), determinando que Random Forest es el más efectivo para predecir la deserción con una precisión del 81.31 %, considerando variables institucionales y estudiantiles.

Tabla 4

Técnicas y algoritmos utilizados en estudios sobre deserción universitaria

Autor	Técnicas/Algoritmos	Mejor precisión
Cedillo et al., 2024	Estadísticas y algebraicas	No especifica
	Redes bayesianas	
	Algoritmos evolutivos	
	Tablas de contingencia Árboles de decisión	
	Redes neuronales artificiales	
Parra et al., 2023	Estocásticas y borrosas	Árboles de decisión
	Arboles de decisión	
	Métodos probabilísticos	
	Regresión logística	
	Redes neuronales	
Cornejo et al., 2023	K-means	RandomTree
	K-vecinos	
	DecisionTable	
	Jrip OneR	
	DecisionStump HoeffdingTree	
Melo, 2024	J48	Random Forest
	LMT	
	RandomForest RandomTree	
	REPTree PART	
	Regresión logística	
Patiño et al., 2024	Árboles de decisión Random Forest	Random Forest
	K-Nearest Neighbors (KNN)	
	Support Vector Machine (SVM)	
	Random Forest	Random Forest
	Naive Bayes LightGBM XGBoost	

Árboles de decisión. Los árboles de decisión, incluyendo algoritmos como C4.5 y Random Forest, han demostrado ser altamente efectivos en la clasificación de estudiantes con riesgo de deserción. Estos modelos permiten crear reglas claras y visuales que segmentan a los estudiantes con base en características como desempeño académico, número de materias reprobadas o asistencia Vélez

et al., (2025). Además, su facilidad de interpretación los convierte en una opción atractiva para instituciones educativas con personal no especializado en análisis de datos.

Redes neuronales. Las redes neuronales son útiles para modelar relaciones complejas entre variables académicas, socioemocionales y de contexto. En estudios recientes, como el de Camargo García (2020), se ha evidenciado su eficacia en la predicción de abandono universitario. Estos modelos destacan por su capacidad de detectar patrones no lineales y ocultos en los datos, aunque presentan desafíos en cuanto a su interpretación (Patiño Hernández et al., 2024).

Máquinas de soporte vectorial (SVM). Las máquinas de soporte vectorial son técnicas robustas para clasificar estudiantes en función de características complejas. Urbina-Nájera et al. (2020) reportan que estas técnicas obtienen altos niveles de precisión al diferenciar entre estudiantes que desertan y los que continúan sus estudios, incluso en contextos de alta dimensionalidad. Sin embargo, su uso requiere un ajuste cuidadoso de parámetros como el *kernel*.

Naive Bayes. El algoritmo Naive Bayes es una opción eficiente y sencilla para la clasificación de estudiantes con alto o bajo riesgo de deserción. Aunque asume independencia entre variables, ha mostrado resultados prometedores en educación superior debido a su bajo costo computacional y su velocidad de entrenamiento Castrillón-Gómez et al., (2020). Por su parte, Patiño Hernández et al. (2024) indican que las técnicas de minería de datos permiten una identificación temprana de los factores que influyen en la deserción estudiantil.

Experimentación

Recopilación de datos y explicación dataset. Para la obtención del dataset, se realizó una búsqueda en el repositorio de la Universidad de California en Irvine (UCI), el cual recopila información para *machine learning*; en este se encontró “Predict Students’ Dropout and Academic Success”, el cual cuenta con información de estudiantes con diferentes datos como: a) el estado civil, b) nacionalidad, c) notas por curso, entre otros. Estos datos están normalizados para facilitar su tratamiento como se muestra en la Figura 2.

Figura 2

Normalización del dataset obtenido de UCI Machine Learning Repository

Variable Name	Role	Type	Demographic	Description	Units	Missing Values
Marital Status	Feature	Integer	Marital Status	1 – single 2 – married 3 – widower 4 – divorced 5 – facto union 6 – legally separated		no

En la Figura 2 se puede ver cómo en la variable Marital Status, la descripción muestra a cada estado civil con un número para así evitar que existan variaciones en la información como, por ejemplo, algún error ortográfico.

Como resultado se dispone del dataset que va a ser utilizado en la experimentación.

Tabla 5
Dataset

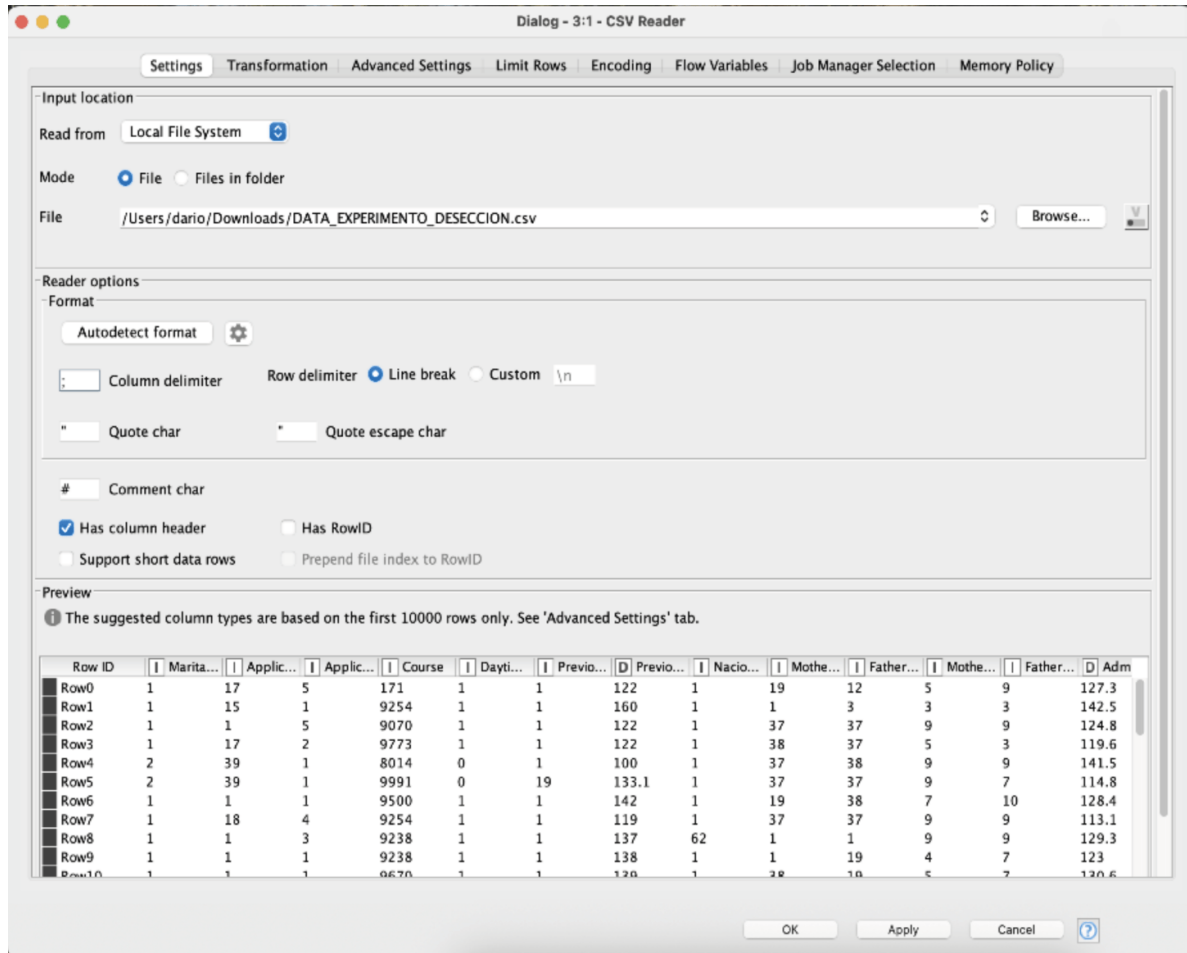
Marital status	Application mode	Application order	...	Target
1	17	5	...	Dropout
1	15	8	...	Graduate
1	1	5	...	Dropout
1	17	2	...	Graduate

Para la experimentación de un modelo de predicción se utilizó el software de analítica de datos KNIME Analytics Platform. La Figura 3 presenta la configuración del nodo CSV Reader en la plataforma KNIME, utilizada para la carga y preprocesamiento de los datos experimentales empleados en la presente investigación. Se cargó el archivo denominado DATA EXPERIMENTO_DESECCION.csv, el cual contiene un conjunto de datos tabulares estructurados, relacionados con un estudio sobre parámetros académicos, socioeconómicos y demográficos de estudiantes universitarios vinculados a eventos de deserción.

El nodo se ha configurado para leer desde el sistema de archivos local, identificando automáticamente el formato del archivo delimitado por comas, el cual reconoce la primera fila como encabezado de columnas. La vista previa en la parte inferior de la interfaz permite verificar los nombres y tipos de las variables como: estado civil, ingresos del solicitante, puntajes académicos previos, nacionalidad, y niveles educativos de los padres. La última columna Adm representa un valor numérico relacionado con el desempeño.

Figura 3

Interfaz de configuración del lector de archivos CSV en KNIME



El conjunto de datos puede representarse como una matriz $X \in R_{n \times m}$, donde n representa el número de observaciones (filas) y m el número de variables (columnas). Cada fila $x_i \in R_m$ corresponde al vector de características de un estudiante i y cada columna x_j representa una variable relevante en el análisis.

Este conjunto de datos constituye la base para posteriores análisis estadísticos, tales como regresión lineal múltiple, clasificación supervisada, o reducción de dimensionalidad mediante análisis de componentes principales (PCA), los cuales permiten modelar la variable dependiente Adm como función de un conjunto de predictores $X_1, X_2, \dots, X_m$:

Donde $y^{\wedge}$ es la estimación del puntaje de desempeño, β_i son los coeficientes de regresión y ϵ es el error aleatorio.

Código fuente para generar el árbol de decisión a través del lenguaje de programación Python

```
# Importación de librerías
import pandas as pd
from sklearn.model_selection import train_test_split
```

```

from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Cargar dataset
df = pd.read_csv("DATA_EXPERIMENTO_DESECCION.csv", sep=";")

# Limpieza de datos
df = df.dropna()

# Separación de variables
X = df.drop(columns=["Target"])
y = df["Target"]
# Codificación de variables categóricas
X = pd.get_dummies(X)

# División entrenamiento / prueba
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=0.2, random_state=42)
# Creación del modelo
clf = DecisionTreeClassifier(max_depth=4, random_state=42)

# Entrenamiento
clf.fit(X_train, y_train)

# Visualización del árbol
plt.figure(figsize=(20, 12), dpi=150)

plot_tree(
    clf,
    feature_names=X.columns,
    class_names=[str(c) for c in clf.classes_],
    filled=True,
    rounded=True,
    fontsize=10,
    impurity=True,
    proportion=True,
    precision=2
)

plt.title("Árbol de Decisión (max_depth=4)", fontsize=14, pad=20)
plt.tight_layout()

# Guardar imagen
plt.savefig("decision_tree.png", dpi=300, bbox_inches="tight")
plt.show()

```

Listing 1: Código Fuente para Generación de Árbol de Decisión

Árbol de decisión generado con Python. El árbol de decisión que se muestra en la Figura 4 tiene una profundidad de cuatro niveles y busca predecir la situación académica de los estudiantes, clasificándolos como desertores, matriculados o graduados. Para ello, se consideran variables relacionadas con el rendimiento en los primeros semestres, el cumplimiento de pagos y factores socioeconómicos como la ocupación de los padres o la edad de ingreso. La variable principal que guía la división del árbol es el número de materias aprobadas en el segundo semestre. Si un estudiante

aprobó con 3.5 o menos, hay una mayor probabilidad de que abandone sus estudios. Este patrón se refuerza cuando también se matriculó en pocas asignaturas o si su madre no tiene empleo formal, lo que refleja el impacto de las condiciones sociales en la continuidad académica.

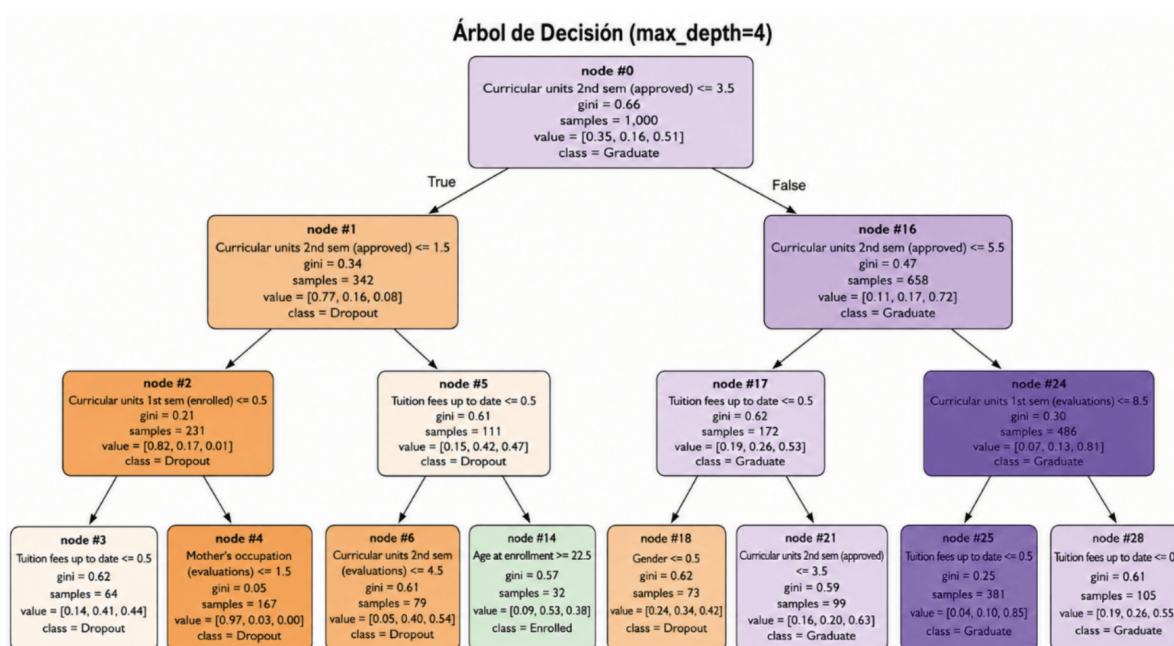
El cumplimiento del pago de matrícula es otro factor relevante. En varios nodos, los atrasos en el pago se asocian con mayores tasas de deserción o con dificultades para graduarse, aunque estos casos presentan índices de Gini elevados, lo que sugiere la influencia de otras variables.

Por otro lado, si un estudiante aprobó más de 3.5 materias y presentó más de 8.5 en las evaluaciones en el primer semestre, tiene una alta probabilidad de que se gradúe. Esto indica que la participación desde el inicio del ciclo académico es un factor clave para el éxito.

Este modelo no solo permite predecir la trayectoria académica de los estudiantes, sino que también ofrece una herramienta útil para identificar factores de riesgo y tomar decisiones institucionales enfocadas en prevenir la deserción.

Figura 4

Árbol de predicción de la situación académica de los estudiantes

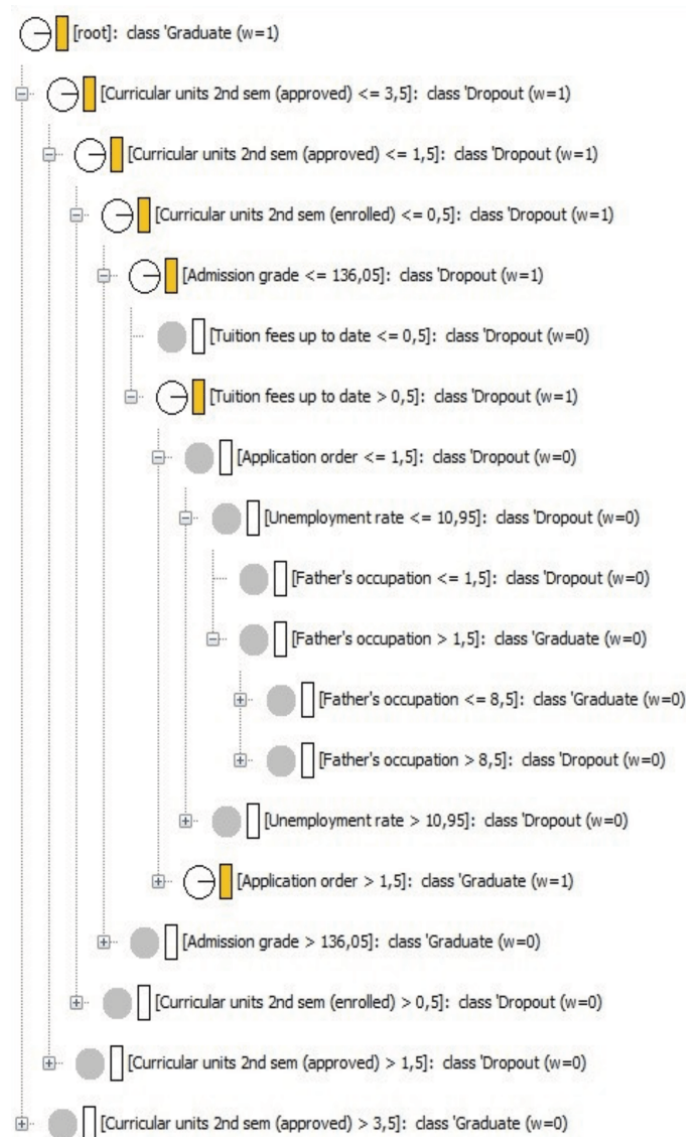


Árbol de decisión, Random Forest. El árbol de decisión se generó en el software KNIME. Para crear este árbol se clasifica a los estudiantes en dos categorías principales: graduados y desertores, utilizando variables académicas, socioeconómicas y administrativas. Se consideran condiciones específicas en cada nodo para segmentar a los estudiantes según perfiles similares. La notación $w = x$ representa el peso de cada regla en la clasificación, siendo mayor cuando la influencia es más significativa. Entre los factores más relevantes destaca el rendimiento académico en el segundo semestre, especialmente el número de materias aprobadas. Aprobar menos de 1.5 o 3.5 asignaturas está fuertemente asociado con la deserción, lo cual evidencia que un bajo desempeño al inicio es un claro indicador de riesgo. El valor de desempeño también cumple un papel importante: puntuaciones menores a 136.05 incrementan la probabilidad de abandono, en línea con investigaciones previas sobre persistencia

estudiantil. El contexto socioeconómico influye considerablemente en los resultados. Por ejemplo, la ocupación del padre tiene un efecto no lineal, donde niveles intermedios se relacionan con mayores tasas de graduación, mientras que los extremos aumentan el riesgo de deserción. De igual manera, una tasa de desempleo regional superior al 10.95 % incrementa la probabilidad de abandono, demostrando el impacto del entorno laboral en la continuidad educativa.

Finalmente, se consideran variables administrativas como el estado de los pagos ya que estudiantes con pagos atrasados tienden a desertar, lo que resalta la importancia de la estabilidad financiera. En conjunto, este modelo permite identificar a tiempo a los estudiantes en riesgo y diseñar estrategias de intervención basadas en datos. Este análisis se evidencia en el árbol de decisión de la Figura 5.

Figura 5
Árbol de decisión generado en KNIME



Por otra parte, la Figura 6 muestra un flujo de trabajo para el análisis de datos mediante un modelo de árbol de decisión. El flujo comienza con la lectura de datos desde dos nodos etiquetados como CSV Reader, los cuales importan datos de entrenamiento y prueba. El nodo “Decision Tree Learner” entrena el modelo que se convierte en reglas con “Decision Tree to Ruleset”. Finalmente, el nodo “Decision Tree Predictor”, aplica el modelo a los datos de prueba, generando predicciones

en una tabla, incluyendo variables como desempleo (10.8), inflación (1.4), PIB (1.74), la variable objetivo y la predicción.

Desde un enfoque matemático, el árbol de decisión minimiza la impureza utilizando el índice de Gini.

El índice de Gini para una región R_m se calcula como:

$$\text{Gini}(R_m) = 1 - \sum_{k=1}^K p_{mk}^2$$

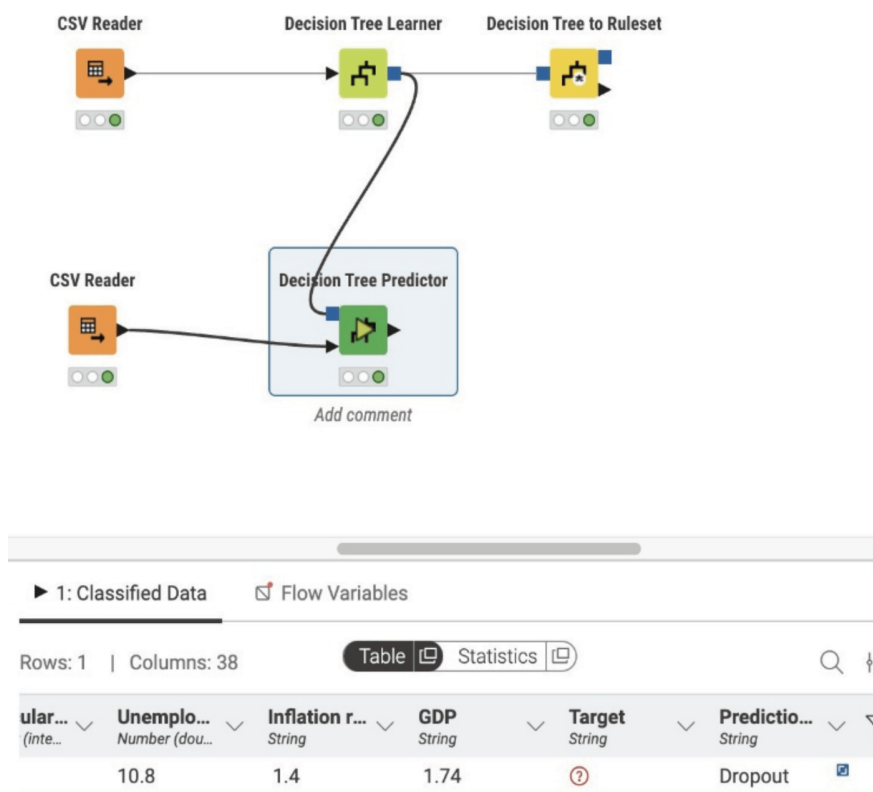
Donde K es el número de clases y p_{mk} es la proporción de observaciones de la clase k en R_m . Un valor bajo indica una región pura, mientras que un valor alto refleja alta impureza. Durante la predicción, para una observación x , el modelo asigna la clase mayoritaria.

$$\hat{y} = \arg \max_k p_{mk} \quad \text{donde } x \in R_m$$

Este flujo destaca un enfoque modular para la clasificación, enfocado en interpretabilidad y análisis económico, permitiendo adaptabilidad para análisis futuros como la evaluación de métricas de desempeño.

Figura 6

Predicción por medio de herramienta de analítica de datos



Código en Python para predecir datos nuevos

Como experimento, se realizó un entorno en KNIME para generar un modelo predictivo. El programa guarda un archivo .pmml, lo cual permite realizar experimentaciones con este modelo de predicción. A continuación, este código en Python se prueba en Colab en el siguiente enlace:

<https://colab.research.google.com/drive/18d8oSUzfSyedFxVoLbAuvEUuTwi71yb#scrollTo=TtDYZ2ONty6s&line=1&uniqifier=1>

```
# Instalación de librería
!pip install -q pypmml

# Importación de librerías
import pandas as pd
from pypmml import Model

# Cargar modelo entrenado
model = Model.load("/content/Prediccion.pmml")

# Cargar datos nuevos
df = pd.read_csv("/content/QUEY_EXPERIMENTO.csv")
# Limpieza de nombres de columnas
df.columns = df.columns.str.replace("'", '').str.replace('\t', '').str.strip()
print("Columnas procesadas:")
print(df.columns.tolist())

# Eliminar variable objetivo si existe
if "Target" in df.columns:
    df = df.drop(columns=["Target"])

# Generar predicciones
predicciones = model.predict(df)

# Agregar resultados
df["Prediccion_Target"] = predicciones["predicted_Target"]
df["Prob_Dropout"] = predicciones["probability_Dropout"]
df["Prob_Graduate"] = predicciones["probability_Graduate"]
df["Prob_Enrolled"] = predicciones["probability_Enrolled"]
print("\nDistribución de clases predichas:")
print(df["Prediccion_Target"].value_counts())
print("\nPredicciones individuales:")
print(df[["Prediccion_Target"]])

# Guardar resultados
df.to_csv("/content/resultados_con_predicciones.csv", index=False)
print("\nArchivo guardado como: resultados_con_predicciones.csv")
```

Listing 2: Código fuente del experimento

A continuación, en la Figura 7 se muestran los resultados de la predicción de abandono estudiantil, implementado mediante un script en Python. En el apartado de “columnas procesadas” se identifican las variables empleadas en el modelo: estado civil, modo de solicitud, orden de

solicitud, curso, asistencia y calificación previa, las cuales representan factores demográficos y académicos relevantes para el análisis.

La sección “Distribución de clases predichas” revela que, para la variable objetivo Prediction Target, se predijo un único caso de abandono (*Dropout*), con un conteo de uno. Este resultado se corrobora en *Predicciones individuales*, donde se confirma que una observación fue clasificada como abandono. Finalmente, los datos procesados se almacenaron en un archivo denominado resultados con predicciones.csv, para su posterior análisis. Este proceso refleja un enfoque sistemático para la identificación de patrones de abandono con énfasis en la estructuración y conservación de los resultados para investigaciones futuras en el ámbito educativo.

Figura 7

Resultados de predicción con Python

```
Columnas procesadas:
['Marital status', 'Application mode', 'Application order', 'Course', 'Daytime/evening attendance', 'Previous qualification']

Distribución de clases predichas:
Prediccion_Target
Dropout      1
Name: count, dtype: int64

Predicciones individuales:
Prediccion_Target
0                Dropout

Archivo guardado como: resultados_con_predicciones.csv
```

Conclusiones

La deserción universitaria en Ecuador es un problema complejo influenciado por múltiples factores. De acuerdo con los estudios revisados, el factor económico es el que tiene mayor incidencia debido a la falta de recursos financieros y la necesidad de trabajar, que afecta especialmente a estudiantes de regiones alejadas y hogares de bajos ingresos. Pero, al desertar del sistema educativo universitario, los empleos a los que tienen acceso las personas son precarios, siendo directamente una consecuencia de haber abandonado sus estudios.

Así mismo, los factores académicos, como el bajo rendimiento y la preparación previa insuficiente, junto con aspectos personales y familiares, como la falta de motivación, problemas emocionales, la ausencia de apoyo y los problemas institucionales, contribuyen significativamente al abandono de los estudios universitarios. Este fenómeno tiene un impacto significativo en las universidades públicas, ya que genera pérdidas económicas por el desperdicio de recursos, afecta su reputación académica, obliga a ajustar procesos internos para fomentar la permanencia estudiantil y restringe las oportunidades de movilidad social de los jóvenes. Es por eso, por lo que el seguimiento institucional es un factor clave para evitar la deserción, el no destinar los recursos suficientes a esto, genera a mediano plazo los anteriormente mencionados problemas institucionales.

Para identificar y predecir casos de abandono de manera temprana, las técnicas de minería de datos han demostrado ser herramientas efectivas. Algoritmos como Random Forest, Random Tree y Árboles de Decisión, destacados por su precisión en estudios recientes, permiten analizar patrones en variables como rendimiento académico, situación socioeconómica y comportamientos estudiantiles. Además, técnicas como redes neuronales, regresión logística, los árboles de decisión y K-Nearest

Neighbors (KNN) ofrecen un enfoque robusto para modelar la deserción, de modo que facilita la implementación de estrategias preventivas por parte de las instituciones educativas. En conjunto, el análisis profundo de los factores de deserción, el uso de técnicas de minería de datos y la adopción de medidas institucionales pueden contribuir significativamente a mejorar la retención estudiantil y el éxito académico en el contexto ecuatoriano.

Referencias

- Acosta, M., Saldaña, C., Olivera, A., Prieto, G., & Guillén, E. (2024). Factores en la deserción estudiantil universitaria en una zona de extrema pobreza del Perú. *Horizontes. Revista de Investigación en Ciencias de la Educación*, 8 (34), 1315-1328. DOI: <https://doi.org/10.33996/revistahorizontes.v8i34.799>
- Amaya, A., Huerta, F., & Flores, C. (2020). Big data, una estrategia para evitar la deserción escolar en las IES. *Revista Iberoamericana de Educación Superior*, 12 (31), 166-178. DOI: <https://doi.org/10.22201/iisue.20072872e.2020.31.712>
- Ávila, L., Cepeda, F., & Aucancela, R. (2024). Deserción en la educación superior en Ecuador, causas y consecuencias. *Ciencia Latina Revista Científica Multidisciplinar*, 8(3), 11475-11490. DOI: https://doi.org/10.37811/cl_rcm.v8i3.12472
- Barahona-Anguisaca, D., Vega-Calvas, P., Moyota-Paguay, A., & Porras-Ramírez, L. (2024). Análisis de modelos estadísticos para predecir el éxito académico en estudiantes universitarios. *MQRInvestigar*, 8(2), 2951-2969. DOI: <https://doi.org/10.56048/MQR20225.8.2.2024.2951-2969>
- Buenaño, E., Beletanga, M. J., y Mancheno, M. (2024). ¿Qué factores son relevantes para entender el abandono? Análisis en una universidad cofinanciada en Ecuador e implicaciones políticas, utilizando modelos de supervivencia de Cox. *J. Latinos Educ.* 23, 1400–1415. <https://doi.org/10.1080/15348431.2023.2271570>
- Camargo, A. (2020). Modelo para la predicción de la deserción de estudiantes de pregrado, basado en técnicas de minería de datos. [Tesis doctoral, Corporación Universidad de la Costa]. Universidad de la Costa. <https://hdl.handle.net/11323/7077>
- Cajigal-Molina, E., Arias-Gómez, L., & Farfan-Heredia, E. (2022). Resiliencia y deserción escolar. Un estudio para plantear estrategias desde la tutoría en la educación superior. *CPU-e, Revista de Investigación Educativa*, (34). DOI: <https://doi.org/10.25009/cpue.v0i34.2794>
- Castrillón, O., Sarache, W., & Ruiz-Herrera, S. (2020). Prediction of academic performance using artificial intelligence techniques. *Formación universitaria*, 13(1), 93-102. <https://dx.doi.org/10.4067/S0718-50062020000100093>
- Castro, M. (2023). *Factores que influyen en la deserción estudiantil en Instituciones de Educación Superior en Latinoamérica durante los años 2017 a 2022: una revisión de literatura*. Universidad Cooperativa de Colombia. <https://hdl.handle.net/20.500.12494/52813>
- Cedillo, J., Beltrán, H., Saltos, M., & Soriano, F. (2024). Explorando la minería de datos en la gestión educativa superior: desafíos y oportunidades en la era digital. *REINCISOL*, 3 (5), 1367-1385. DOI: [https://doi.org/10.59282/reincisol.V3\(5\)1367-1385](https://doi.org/10.59282/reincisol.V3(5)1367-1385)

- Conde, S., García, M., & Toscano, M. (2023). Riesgo de abandono escolar: ¿Cómo influyen las características sociofamiliares percibidas por los estudiantes sobre sus actitudes y comportamiento en el aula? *Educación XXI*, 26 (2), 267-298. <https://doi.org/10.5944/educxx1.33279>
- Cornejo, M., Vega, L., Naranjo, M., Acosta, O., Ávila, F., & Sotomayor, M. de la A. (2023). Modelo predictivo de la deserción escolar en educación superior: una aproximación desde la minería de datos utilizando la metodología CRISP-DM. *Ciencia Latina Revista Científica Multidisciplinar*, 7 (5), 7797-7812. DOI: https://doi.org/10.37811/cl_rcm.v7i5.8363
- De Gracia Vega, J. Y., & Vega de Martínez, I. (2024). Factores que contribuyen a la deserción de los estudiantes que cursan la licenciatura en educación primaria, del anexo Cerro Puerco, Comarca Ngäbe Buglé, 2022. *Revista Colegiada de Ciencia*. DOI: <https://doi.org/10.48204/j.colegiada.v6n1.a5875>
- Delgado Saeteros, Z. (2024). Análisis de la deserción estudiantil y estrategias para incrementar la retención en instituciones de educación superior. *Revista de Investigación, Formación y Desarrollo: Generando Productividad Institucional*, 12 (1), 88-95. DOI: <https://doi.org/10.34070/rif.v12.i1.WVVF7996>
- Duque, R., Araujo, I., Cadena-Palacios C., & Fuertes, B. (2024). Importancia de la orientación profesional previa al acceso a la enseñanza superior. *Revista Científica UISRAEL*, 11(1), 13-30. <https://doi.org/10.35290/rcui.v11n1.2023.1058>
- Escalante, J., Medina, C., & Vásquez, A. (2023). *La deserción universitaria: un problema no resuelto en el Perú*. HACEDOR-AIAPAEC DOI: <https://doi.org/10.26495/rch.v7i1.2421>
- Fonseca-Grandón, G., Gabalán-Coello, J., & Vásquez-Rizo, F. (2022). Tipologías de estudiantes que permanecen y abandonan la Universidad: aproximación comprensiva desde el currículum y el proyecto de vida. *Revista Virtual Universidad Católica del Norte*, (65), 5-30. <https://www.doi.org/10.35575/rvucn.n65a2>
- Galve, C. (2024). *Estudio de las variables influyentes en el abandono universitario* (Doctoral dissertation). Universidad de Oviedo. <https://digibuo.uniovi.es/dspace/handle/10651/75635>
- Granda, L. E. Á., et al. (2024). *Deserción en la educación superior en Ecuador, causas y consecuencias*. Ciencia Latina Revista Científica Multidisciplinar. <https://dialnet.unirioja.es/servlet/articulo?codigo=9783947>
- Guadalupe, V. (2025). *Modelo predictivo basado en machine Learning para la reducción de la deserción estudiantil en las universidades privadas del Perú: caso Universidad Privada San Juan Bautista*. Universidad Nacional Federico Villarreal. <https://hdl.handle.net/20.500.13084/10839>
- Gutiérrez, R., Nuñez, H., & López, E. (2024). Reflexiones teóricas del fenómeno de la deserción académica en la educación superior. *Revie-Revista de Investigación y Evaluación Educativa*, 11(2), 88-109. <https://doi.org/10.47554/revie.vol11.num2.2024.pp88-109>

- Herreño-Munera, M., Romero-Tenorio, J., Mejía-Ríos, J., & Román-Santana, W. (2024). Deserción estudiantil en educación superior. Tendencias y oportunidades en la era post pandemia. *Revista Arbitrada Interdisciplinaria Koinonía*, 9(18), 156-177. Epub 15 de noviembre de 2024. <https://doi.org/10.35381/r.k.v9i18.4201>
- Instituto Nacional de Estadística y Censos del Ecuador (INEC). (2024). *Encuesta Nacional de Empleo, Desempleo y Subempleo*. <https://www.ecuadorencifras.gob.ec>
- Jarrín, G., & Moreta-Herrera, R. (2024). Estrés, Dificultades de Regulación Emocional y Adaptación Escolar en adolescentes aspirantes a la educación superior en Ecuador. *ReHuSo: Revista de Ciencias Humanísticas y Sociales*. DOI: <https://doi.org/10.33936/rehuso.v9i1.5772>
- Katt, L., Xicot'encatl-Ramírez, G., & Ibarra-Corona, M. (2024). Factores que influyen en el abandono escolar en la Universidad Tecnológica del Sur- este de Veracruz, México. *Transdigital*, 5 (10), e347. <https://doi.org/10.56162/transdigital347>
- Loaiza, D., García, I., Romero, J., Díaz, M., & Ronquillo, P. (2023). Identificación de los factores de la deserción académica en el sistema educativo del Ecuador. *Ciencia Latina Revista Científica Multidisciplinar*, 7 (2), 11121-11136. https://doi.org/10.37811/cl_rcm.v7i2.6190
- López, J. (2023). *Deserción universitaria y su relación con las afectaciones psicológicas*. Universidad Cooperativa de Colombia, Facultad de Ciencias Sociales, Psicología, Bogotá. Disponible en: <https://hdl.handle.net/20.500.12494/53283>
- Luna-Rizo, M., Daza, M., Lozoya, J., Rocha, A., Madrigal, R., Ramos, B., Cortés, L., González, A., Arámburo, J., García, J., Cisneros, M., Ramírez, M., Beltrán, J., Arce, M. de J., Rodríguez, D., López, J., Llanes, C., Maciel, M. del R., González, B., ... Aréchiga Cabrera, V. D. (2024). *Tendencias de la inteligencia artificial en educación (1.ª ed.)*. Editorial Universidad de Guadalajara. https://mta.udg.mx/sites/default/files/adjuntos/tendencias_de_la_inteligencia_artificial.pdf
- Mazurett, M. (2025). Modelos predictivos de deserción estudiantil incorporando variables de salud mental en la Universidad de Chile para elaborar estrategias de intervención temprana. Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/205220>
- Melo, M. (2024). Modelo predictivo de los factores de deserción escolar de las instituciones educativas en bachillerato en el Ecuador de los años 2012–2021 [Tesis de maestría]. Pontificia Universidad Católica del Ecuador. <https://repositorio.puce.edu.ec/handle/123456789/45106>
- Ministerio de Educación Nacional. (2022). Estrategias para promover la permanencia en educación superior. <https://www.mineduacion.gov.co/portal/>
- Moreta-Herrera, R., Zambrano-Estrella, J., Sánchez-Vélez, H., & Naranjo-Vaca, S. (2021). Salud mental en universitarios del Ecuador: síntomas relevantes, diferencias por género y prevalencia de casos. *Pensamiento Psicológico*, 19(1). <https://doi.org/10.11144/Javerianacali.PPSI19.smue>

- Muñoz, R., & Cedeño, T. (2024). La salud mental en la universidad: Soledad no deseada en universitarios ecuatorianos: causas y consecuencias. *Dialoguemos*. <https://dialoguemos.ec/2024/10/la-salud-mental-en-la-universidad-la-soledad-no-deseada-despunta-entre-los-estudiantes>
- Muñoz, F. (2024). Evaluación de modelo predictivo del abandono universitario basado en clusters de comportamiento de los/as estudiantes de primer año de ingeniería udec [Tesis Doctoral]. Universidad de Concepción. <https://repositorio.udec.cl/server/api/core/bitstreams/cfea2a6b-ff78-4222-98da-615abadc3a34/content>
- Noriega, E., Becerra-Sánchez, A., Zepeda-Valles, G., & González-Arroyo, R. (2024). *Minería de datos para la predicción del rendimiento académico de estudiantes universitarios. Memorias del Congreso Internacional de Mecatrónica, Control e Inteligencia Artificial (Vol. 3)*. Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Cuautitlán. <https://virtual.cuautitlan.unam.mx/intar/memoriascimcia/wp-content/uploads/sites/20/2025/09/24-Mineria-de-datos-1.pdf>
- Núñez-Naranjo, A. F. (2024). Analysis of the determinant factors in university dropout: a case study of Ecuador. *Frontiers in Education*, 9, 1444534. DOI: <https://doi.org/10.3389/feduc.2024.1444534>
- Orellana-Correa, G., Ayala, J., & Vega, A. (2024). Factores socioeconómicos que inciden en la deserción universitaria, en la carrera de Economía, Universidad Técnica de Machala, Facultad de Ciencias Empresariales. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, 5 (4), 2788-2801. <https://doi.org/10.56712/latam.v5i4.2455>
- Osorio-González, J. (2024). Comparación de modelos estadísticos y de aprendizaje automático para la predicción y prevención de la deserción estudiantil en los programas de pregrado de la Universidad EAFIT. Universidad EAFIT. <https://hdl.handle.net/10784/34646>
- Patiño, L., Reina, A., Sawicki, S., Frantz, R., & Roos-Frantz, F. (2024). Predicción de desempeño académico basado en análisis de datos usando técnicas de clasificación: un estudio con datos de una universidad pública de Ecuador. *Revista Novas Tecnologías na Educação*, 22 (1), 497-507. <https://seer.ufrgs.br/index.php/renote/article/view/141575> <http://gca.unijui.edu.br/publication/ae0e08163d22befd4635f47bef1b6e3f.pdf>
- Parra-Sánchez, J., Torres, I., & Martínez, C. (2023). Factores explicativos de la deserción universitaria abordados mediante inteligencia artificial. *Revista Electrónica de Investigación Educativa*, 25 (e18), 1-17. DOI: <https://doi.org/10.24320/redie.2023.25.e18.4455>
- Parraga-Menchaca, D. (2024). Modelos Predictivos de Rendimiento Académico Universitario Mediante Aprendizaje Automático. *Revista Científica De Salud Y Desarrollo Humano*, 5(2), 974–991. DOI: <https://doi.org/10.61368/r.s.d.h.v5i2.204>

- Passeggia, R., Testa, I., Esposito, G., Picione, R., Ragozini, G., & Freda, M. (2023). Examining the Relation Between First-year University Students' Intention to Drop-out and Academic Engagement: The Role of Motivation, Subjective Well-being and Retrospective Judgements of School Experience. *Innov High Educ*, 48, 837–859. <https://doi.org/10.1007/s10755-023-09674-5>
- Peet, T. (2025). Análisis de patrones de comportamiento estudiantil en aprendizaje en línea para la caracterización de cursos [Tesis de pregrado]. Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/205209>
- Quincho, R., Carrillo, J., Ccencho, A., Inga, V., Cárdenas, J., & Huamán, D. (2024). University Dropout: A Systematic Review of the Main Determinant Factors (2020-2024). *F1000Research*, 13, 942. <https://doi.org/10.12688/f1000research.154263.2>
- Rabelo, A., & Zárate, L. (2024). A model for predicting dropout of higher education students. *Data Science and Management*, 8(1), 72-85. DOI: <https://doi.org/10.1016/j.dsm.2024.07.001>
- Rebelo, M., Reis, T., Thompsen, T. Targino, R., Ramos, V., Marques Queiroga, E., Munoz, R., & Cechinel, C. (2025). Student dropout prediction through machine learning optimization: insights from moodle log data. *Scientific Reports* 15(1), 9840. <https://doi.org/10.1038/s41598-025-93918-1>
- Salcedo-Escarria, A. (2020). Deserción universitaria en Colombia. *Academia y Virtualidad*, 3 (1), 50-60. <https://revistas.umng.edu.co/index.php/ravi/article/view/5461>
- Salinas, M. (2023). Crisis económica post pandemia: cómo afecta en el acceso a la educación superior en el Paraguay. Caso: universidades públicas de Ayolas. *Ciencia Latina Revista Científica Multidisciplinar*, 7 (4), 6496-6506. https://doi.org/10.37811/cl_rcm.v7i4.7422
- Terrazas, A., Velázquez-Castro, J., & Testón-Franco, N. (2022). El estrés académico y afectaciones emocionales en estudiantes de nivel superior. *Revista Innova Educación*, 4 (2), 132-146. <https://dialnet.unirioja.es/descarga/articulo/8326422.pdf>
- Ujkani, B., Minkovska, D., & Stoyanova, L. (2022). Application of Logistic Regression Technique for Predicting Student Dropout. 2022 XXXI *International Scientific Conference Electronics (ET)*, pp. 1-4. DOI: <https://ieeexplore.ieee.org/document/9920280>
- Urbina-Nájera, A., Camino-Hampshire, J., & Cruz, R. (2020). Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *RELIEVE—Revista Electrónica De Investigación Y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.16061>
- Vélez, L., Chávez, J., Berruz, M., Vite, Y., & Ruiz, M. (2025). Factores externos que influyen en la deserción universitaria en el Ecuador. *Revista Latinoamericana de Calidad Educativa*, 2 (1), 299-308.
- Vidal, J. (2023). Factores motivacionales, emocionales y socioeconómicos asociados al rendimiento académico y abandono en la Educación Superior: Estudio en dos Universidades del Ecuador [Tesis doctoral]. Universidad de Alicante. <https://hdl.handle.net/10045/132689>

Villegas, C., & Núñez, D. (2024). Estrategias institucionales para prevenir la deserción en universidades ecuatorianas. *Revista de Educación Superior*, 12 (1), 40-60. <https://www.ride.org.mx/index.php/RIDE/article/view/1923>