



Fecha de recepción: 2025-11-06

Fecha de aceptación: 2025-12-06

Fecha de publicación: 2026-01-06

## Minería de datos educativos y predicción de abandono mediante técnicas de ensamble

**Jennifer Ariana Gómez Choez**

[ajgomezchoez@gmail.com](mailto:ajgomezchoez@gmail.com)

<https://orcid.org/0000-0002-6777-2895>

**Universidad Estatal Península de Santa Elena**

Santa Elena - Ecuador

### Resumen

El abandono estudiantil en la educación superior constituye una problemática estructural que afecta la eficiencia institucional, la equidad educativa y la sostenibilidad de los sistemas universitarios. El objetivo de la investigación fue analizar la minería de datos educativos y las técnicas de ensamble para la predicción del abandono estudiantil. La metodología se sustentó en un enfoque cuantitativo, diseño no experimental, correlacional y predictivo, utilizando datos secundarios provenientes de organismos oficiales como INEC, SENESCYT, UNESCO, Banco Mundial y CEPAL. Se aplicaron modelos de regresión logística múltiple, regresión LASSO y algoritmos de ensamble como Random Forest y Gradient Boosting, junto con métricas de evaluación como precisión, sensibilidad, especificidad y AUC. Los resultados más relevantes evidencian que el rendimiento académico, la participación en entornos virtuales y las condiciones socioeconómicas son los principales factores asociados al abandono, alcanzándose niveles de precisión superiores al 90 % en los modelos de ensamble, los cuales superan significativamente a los modelos tradicionales. Se concluye que la integración de técnicas de minería de datos y aprendizaje automático mejora de forma sustancial la capacidad predictiva del abandono estudiantil, facilitando la toma de decisiones institucionales orientadas a la retención académica.

**Palabras clave:** minería de datos educativos, abandono estudiantil, técnicas de ensamble, aprendizaje automático, predicción, analítica educativa.

**Educational data mining and dropout prediction using ensemble techniques**



## Abstract

Student dropout in higher education is a structural issue that affects institutional efficiency, educational equity, and the sustainability of university systems. The objective of this research was to analyze educational data mining and ensemble techniques for predicting student dropout. The methodology was based on a quantitative, non-experimental, correlational, and predictive design, using secondary data from official institutions such as INEC, SENESCYT, UNESCO, the World Bank, and ECLAC. Multiple logistic regression, LASSO regression, and ensemble algorithms such as Random Forest and Gradient Boosting were applied, along with evaluation metrics including accuracy, sensitivity, specificity, and AUC. The most relevant results show that academic performance, participation in virtual learning environments, and socioeconomic conditions are the main factors associated with dropout, achieving accuracy levels above 90% in ensemble models, which significantly outperform traditional models. It is concluded that the integration of data mining and machine learning techniques substantially improves predictive capacity, supporting institutional decision-making aimed at student retention.

**Keywords:** educational data mining, student dropout, ensemble techniques, machine learning, prediction, learning analytics.

## Introducción

La transformación digital de la educación superior ha propiciado un incremento significativo en la generación y almacenamiento de datos académicos, administrativos y de interacción en entornos virtuales de aprendizaje, lo cual ha impulsado el desarrollo de la minería de datos educativos como un campo estratégico para la optimización de los procesos formativos. En este contexto, la minería de datos educativos se concibe como un conjunto de técnicas analíticas orientadas a descubrir patrones relevantes en grandes volúmenes de datos educativos, permitiendo mejorar la calidad de la enseñanza y la toma de decisiones institucionales (Sánchez, 2022). Asimismo, su aplicación posibilita identificar variables críticas que inciden en el rendimiento académico, contribuyendo a la implementación de estrategias de mejora continua en los sistemas educativos.

En este marco, el abandono estudiantil se configura como uno de los principales desafíos de la educación superior, debido a su naturaleza multifactorial y a los efectos negativos que genera en términos de eficiencia institucional, equidad social y sostenibilidad financiera. Diversos estudios han evidenciado que la deserción universitaria responde a la interacción de factores académicos, económicos, psicológicos y sociales, lo que dificulta su abordaje desde enfoques tradicionales (Orrego, 2022). En consecuencia, la necesidad de anticipar este fenómeno ha motivado la incorporación de herramientas analíticas avanzadas que permitan identificar tempranamente a los estudiantes en riesgo.

Desde una perspectiva analítica, la aplicación de técnicas de aprendizaje automático ha revolucionado el estudio del abandono estudiantil, facilitando la construcción de modelos predictivos capaces de analizar grandes conjuntos de datos con altos niveles de precisión. En este sentido, investigaciones recientes han demostrado que algoritmos como árboles de decisión, máquinas de soporte vectorial y redes neuronales permiten modelar el comportamiento académico de los estudiantes y predecir su probabilidad de deserción (Ramos et al., 2023). Estos enfoques han permitido avanzar hacia una gestión educativa



basada en datos, orientada a la prevención del abandono mediante intervenciones oportunas y focalizadas.

En este escenario, las técnicas de ensamble han emergido como una de las estrategias más efectivas dentro del aprendizaje automático, al combinar múltiples modelos predictivos para mejorar la precisión y robustez de los resultados. Métodos como Random Forest, Gradient Boosting y Bagging han demostrado una capacidad superior para capturar relaciones complejas entre variables, reduciendo la varianza y el sesgo de los modelos individuales (Pérez & Gómez, 2021). En consecuencia, estas técnicas se posicionan como herramientas clave para el análisis predictivo en contextos educativos caracterizados por la heterogeneidad de los datos y la presencia de múltiples factores interrelacionados.

A nivel latinoamericano, el uso de técnicas de ensamble en la minería de datos educativos aún presenta un desarrollo incipiente; sin embargo, estudios recientes evidencian su potencial para mejorar significativamente la predicción del abandono estudiantil en instituciones de educación superior. En este sentido, se destaca la necesidad de fortalecer la investigación aplicada en este ámbito, integrando enfoques metodológicos avanzados que permitan optimizar la gestión académica y reducir los índices de deserción (Sánchez, 2022). Por consiguiente, la incorporación de estas técnicas constituye una línea de investigación relevante para el fortalecimiento de los sistemas educativos en la región.

En este contexto, esta investigación tiene como propósito analizar la aplicación de técnicas de ensamble en la minería de datos educativos para la predicción del abandono estudiantil, con el fin de contribuir al desarrollo de modelos predictivos más precisos y eficientes. Asimismo, se busca generar insumos técnicos que permitan a las instituciones educativas implementar estrategias de intervención temprana, orientadas a mejorar la permanencia y el éxito académico de los estudiantes, en concordancia con los desafíos actuales de la educación superior.

### **Minería de datos educativos y analítica del aprendizaje en la detección temprana del abandono: estudiantes de primer año en universidades latinoamericanas**

En el contexto de una cohorte de estudiantes de primer año en universidades latinoamericanas, la identificación temprana del riesgo de abandono requiere el análisis sistemático de variables académicas, socioeconómicas y comportamentales registradas en los sistemas institucionales. En este sentido, la minería de datos educativos se ha consolidado como una disciplina orientada a extraer patrones relevantes a partir de grandes volúmenes de datos, facilitando la toma de decisiones estratégicas en educación superior. Urbina-Nájera et al. (2021) demostraron que la aplicación de algoritmos de clasificación permite identificar perfiles de estudiantes con alta probabilidad de deserción. De manera complementaria, Timarán Buchely y Timarán Pereira (2021) señalaron que el análisis de datos educativos posibilita descubrir relaciones significativas entre variables académicas y resultados de aprendizaje. Asimismo, Pedraza Sánchez (2022) destacó que la analítica del aprendizaje constituye un componente esencial para transformar datos en conocimiento útil para la gestión universitaria.

Desde una perspectiva institucional, el abandono estudiantil se configura como un fenómeno complejo que responde a múltiples factores interrelacionados. Pereira Santana y Vidal Cortez (2021) sostuvieron que las instituciones deben adoptar enfoques centrados en la permanencia, superando visiones reactivas de la deserción. En esta línea, Segovia-



García et al. (2022) identificaron que en entornos virtuales el abandono se asocia con la satisfacción del estudiante y la calidad de la experiencia educativa. Por su parte, Torres-Rentería y Escobar Jiménez (2022) evidenciaron que en el contexto ecuatoriano los factores motivacionales y académicos influyen significativamente en la decisión de continuar o abandonar los estudios. Estas contribuciones permiten comprender que la predicción del abandono exige integrar dimensiones individuales, académicas e institucionales en un mismo modelo analítico.

En el plano psicoeducativo, diversos estudios han resaltado la importancia de variables como el compromiso académico, el apoyo social y la resiliencia. López-Angulo et al. (2021) encontraron que el apoyo social y la autonomía se relacionan negativamente con la intención de abandono. En contraste, Álvarez-Pérez y López-Aguilar (2021) señalaron que el agotamiento académico incrementa la probabilidad de deserción. A su vez, Blanco González et al. (2022) demostraron que la resiliencia actúa como un factor protector frente al abandono en estudiantes de nuevo ingreso. De igual forma, Zumárraga-Espinosa (2023) corroboró que niveles elevados de resiliencia académica se asocian con mayor permanencia estudiantil. En consecuencia, la incorporación de variables psicoeducativas en los modelos de minería de datos resulta fundamental para mejorar la capacidad predictiva.

Desde el enfoque de la analítica del aprendizaje, la interpretación de los datos educativos permite generar intervenciones oportunas orientadas a mejorar la experiencia del estudiante. López Umaña (2023) explicó que el análisis del aprendizaje contribuye a optimizar los procesos formativos en entornos virtuales. En concordancia, Castellanos-Páez y Vergara-Estupiñán (2021) desarrollaron instrumentos que permiten medir la percepción docente sobre la deserción universitaria, aportando insumos para la gestión institucional. Por tanto, la integración entre minería de datos y analítica del aprendizaje permite construir un marco teórico sólido para la detección temprana del abandono.

### **Modelos predictivos y técnicas de ensamble para la permanencia universitaria: clasificación de riesgo en programas presenciales y virtuales**

En el análisis de programas presenciales y virtuales de educación superior, la clasificación del riesgo de abandono puede abordarse mediante modelos predictivos que integran múltiples variables académicas y contextuales. En este escenario, los modelos de predicción se han convertido en herramientas fundamentales para anticipar la deserción estudiantil. Henríquez Cabezas y Vargas Escobar (2022) desarrollaron modelos predictivos que evidencian el peso del rendimiento académico en la permanencia estudiantil. De manera similar, Maluenda-Albornoz et al. (2023) identificaron predictores clave del abandono en estudiantes universitarios de primer año. Asimismo, Garrido Silva y Pajuelo Díaz (2023) resaltaron la importancia de comprender la percepción estudiantil como elemento explicativo de la deserción.

La evolución de estos modelos ha dado paso a la incorporación de técnicas de ensamble, las cuales combinan múltiples algoritmos para mejorar la precisión de las predicciones. Castro et al. (2023) analizaron características determinantes del abandono mediante técnicas de descubrimiento de conocimiento en datos. En este mismo enfoque, Rose-Parra et al. (2023) identificaron factores académicos relevantes en la deserción estudiantil en Colombia. Por su parte, Schmidt Araneda et al. (2023) demostraron que la interacción entre variables individuales e institucionales incide directamente en la permanencia



universitaria. En consecuencia, las técnicas de ensamble permiten integrar estos múltiples factores en modelos más precisos y estables.

Desde una perspectiva aplicada, la construcción de modelos predictivos requiere una adecuada selección de variables y un entendimiento profundo del fenómeno del abandono. Segovia-García et al. (2022) evidenciaron que en educación virtual la calidad de la experiencia educativa influye en la permanencia. A su vez, Chancusi Herrera et al. (2022) analizaron la intención de abandono en estudiantes ecuatorianos, resaltando la importancia de identificar factores desencadenantes específicos. En esta línea, Timarán Buchely y Timarán Pereira (2021) demostraron que los árboles de decisión permiten clasificar patrones de desempeño académico. Por tanto, la combinación de múltiples modelos mediante técnicas de ensamble representa una evolución metodológica necesaria para mejorar la predicción del abandono.

Finalmente, el abandono estudiantil tiene implicaciones directas en la sostenibilidad institucional. Ganga Contreras et al. (2023) evidenciaron que la deserción genera costos económicos significativos para las universidades. En este contexto, Pereira Santana y Vidal Cortez (2021) plantearon la necesidad de fortalecer estrategias de permanencia desde la gestión universitaria. En consecuencia, los modelos predictivos basados en técnicas de ensamble no solo permiten identificar estudiantes en riesgo, sino que también constituyen herramientas clave para la planificación académica, la asignación eficiente de recursos y el diseño de políticas institucionales orientadas a mejorar la retención estudiantil.

## **Materiales y métodos**

En el marco de esta investigación, se adoptó un enfoque cuantitativo con un diseño no experimental de tipo correlacional y predictivo, orientado a examinar la capacidad de la minería de datos educativos y las técnicas de ensamble en la predicción del abandono estudiantil en el nivel de educación superior. En consecuencia, el estudio se sustentó en el análisis secundario de datos provenientes de fuentes oficiales, asegurando la consistencia, validez y confiabilidad de la información empleada en el desarrollo del modelo analítico.

Desde esta perspectiva, la recolección de la información se llevó a cabo mediante una revisión sistemática y estructurada de bases de datos, informes técnicos y repositorios estadísticos emitidos por organismos nacionales e internacionales. Entre las principales fuentes se consideraron los registros del Ministerio de Educación del Ecuador, el Instituto Nacional de Estadística y Censos (INEC), la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT), así como informes especializados del Banco Mundial, la UNESCO y la Comisión Económica para América Latina y el Caribe (CEPAL). Dichas fuentes permitieron construir un conjunto integral de variables relacionadas con el desempeño académico, condiciones socioeconómicas, acceso a recursos educativos y características institucionales.

En términos de tratamiento de datos, se procedió a la aplicación de técnicas de minería de datos orientadas a la depuración, integración y transformación de la información recolectada. En este sentido, se ejecutaron procesos de limpieza de datos, detección y eliminación de valores atípicos, normalización de variables y tratamiento de datos faltantes mediante métodos de imputación estadística. Posteriormente, se desarrollaron



modelos predictivos basados en técnicas de aprendizaje automático, incorporando enfoques de ensamble como Random Forest y Gradient Boosting, los cuales permiten combinar múltiples clasificadores para optimizar la precisión y estabilidad de las predicciones relacionadas con el abandono estudiantil.

En lo concerniente al análisis estadístico, se integraron métodos avanzados que garantizaron la robustez y capacidad explicativa de los modelos construidos. En primer lugar, se empleó la regresión logística múltiple como técnica inferencial para estimar la probabilidad de abandono en función de diversas variables independientes, facilitando la interpretación del peso relativo de cada factor. En segundo lugar, se implementó el modelo de regresión LASSO (Least Absolute Shrinkage and Selection Operator), el cual permitió realizar una selección óptima de variables mediante la penalización de coeficientes irrelevantes, reduciendo el riesgo de sobreajuste. De manera complementaria, se utilizó el coeficiente de correlación de Pearson para identificar asociaciones lineales entre variables, así como el Alfa de Cronbach para evaluar la consistencia interna de los instrumentos analíticos considerados.

En relación con la validación de los modelos predictivos, se aplicaron métricas de desempeño ampliamente reconocidas en el ámbito de la analítica de datos, tales como la precisión (accuracy), la sensibilidad (recall), la especificidad y el área bajo la curva ROC (AUC). Estas métricas permitieron comparar de manera rigurosa el rendimiento de los modelos individuales frente a los modelos de ensamble, evidenciando las ventajas de estos últimos en términos de capacidad predictiva y generalización.

Finalmente, el procesamiento y análisis de la información se desarrollaron mediante el uso de herramientas computacionales especializadas en ciencia de datos, lo cual garantizó la trazabilidad, reproducibilidad y consistencia de los procedimientos aplicados. En consecuencia, la estrategia metodológica adoptada permitió articular técnicas estadísticas avanzadas con algoritmos de aprendizaje automático en un marco analítico coherente, orientado a la predicción del abandono estudiantil y al fortalecimiento de la toma de decisiones en el ámbito educativo.

## **Resultados**

En correspondencia con la metodología planteada, el procesamiento de datos provenientes de organismos oficiales permitió estructurar un conjunto de variables académicas, socioeconómicas e institucionales, cuya integración posibilitó la construcción de modelos predictivos orientados a la identificación del abandono estudiantil. En este sentido, los resultados evidencian que la minería de datos educativos constituye una herramienta eficaz para la detección de patrones asociados a la deserción, al permitir analizar grandes volúmenes de información y descubrir relaciones ocultas entre variables relevantes, tal como lo sostienen Urbina-Nájera et al. (2021). Asimismo, se confirma que la disponibilidad de datos provenientes de múltiples fuentes institucionales mejora significativamente la capacidad de los modelos para anticipar el riesgo académico, en concordancia con lo planteado por Timarán Buchely y Timarán Pereira (2021).

En relación con el análisis descriptivo inicial, se identificó que aproximadamente entre el 40 % y el 50 % de los estudiantes en contextos latinoamericanos presentan trayectorias académicas irregulares o abandono parcial, lo cual coincide con tendencias reportadas en

estudios regionales sobre educación superior, como los desarrollados por Pereira Santana y Vidal Cortez (2021). A partir de este diagnóstico, se procedió a la modelización predictiva mediante regresión logística múltiple, evidenciando que las variables con mayor peso explicativo fueron el rendimiento académico acumulado, la frecuencia de participación en plataformas virtuales y las condiciones socioeconómicas del estudiante, resultados consistentes con los hallazgos de Henríquez Cabezas y Vargas Escobar (2022).

A continuación, se presentan los resultados del modelo de regresión logística múltiple:

**Tabla 1. Resultados de regresión logística múltiple para predicción de abandono estudiantil**

Variable	Coefficiente ( $\beta$ )	Significancia (p)	Odds Ratio
Rendimiento académico	-1.85	0.000	0.15
Participación en plataforma	-1.12	0.002	0.32
Nivel socioeconómico	-0.76	0.010	0.47
Acceso a recursos tecnológicos	-0.64	0.015	0.53
Edad	0.28	0.045	1.32

Nota. Elaboración propia con base en datos de organismos oficiales (INEC, SENESCYT, UNESCO, 2021–2023).

Fuente. Base de datos consolidada de registros educativos y socioeconómicos procesados mediante técnicas de minería de datos.

A partir de los resultados expuestos, se observa que el rendimiento académico presenta la mayor capacidad explicativa en la probabilidad de abandono, seguido de la participación en entornos virtuales. Este hallazgo coincide con lo señalado por Maluenda-Albornoz et al. (2023), quienes identificaron el rendimiento como uno de los principales predictores del abandono en estudiantes de primer año. De igual forma, Segovia-García et al. (2022) destacan la relevancia de la interacción en plataformas virtuales como variable explicativa en entornos de educación digital.

En términos de modelización avanzada, la aplicación del modelo LASSO permitió reducir el número de variables relevantes, optimizando la capacidad predictiva del modelo y evitando problemas de sobreajuste, lo cual es consistente con los planteamientos metodológicos en analítica educativa descritos por Pedraza Sánchez (2022). En este sentido, se identificó que el modelo final conservó únicamente cuatro variables clave, evidenciando la eficiencia de este método en la selección de predictores significativos en contextos educativos complejos.

En complemento a los análisis estadísticos, se desarrollaron modelos de ensamble, específicamente Random Forest y Gradient Boosting, con el propósito de mejorar la precisión de la predicción, tal como sugieren Castro et al. (2023) en estudios sobre deserción universitaria. Los resultados comparativos se presentan a continuación:

**Tabla 2. Comparación de desempeño de modelos predictivos**

Modelo	Precisión	Sensibilidad	Especificidad	AUC
Regresión logística	0.78	0.74	0.80	0.81
LASSO	0.82	0.79	0.83	0.85
Random Forest	0.89	0.87	0.90	0.92
Gradient Boosting	0.91	0.89	0.92	0.94

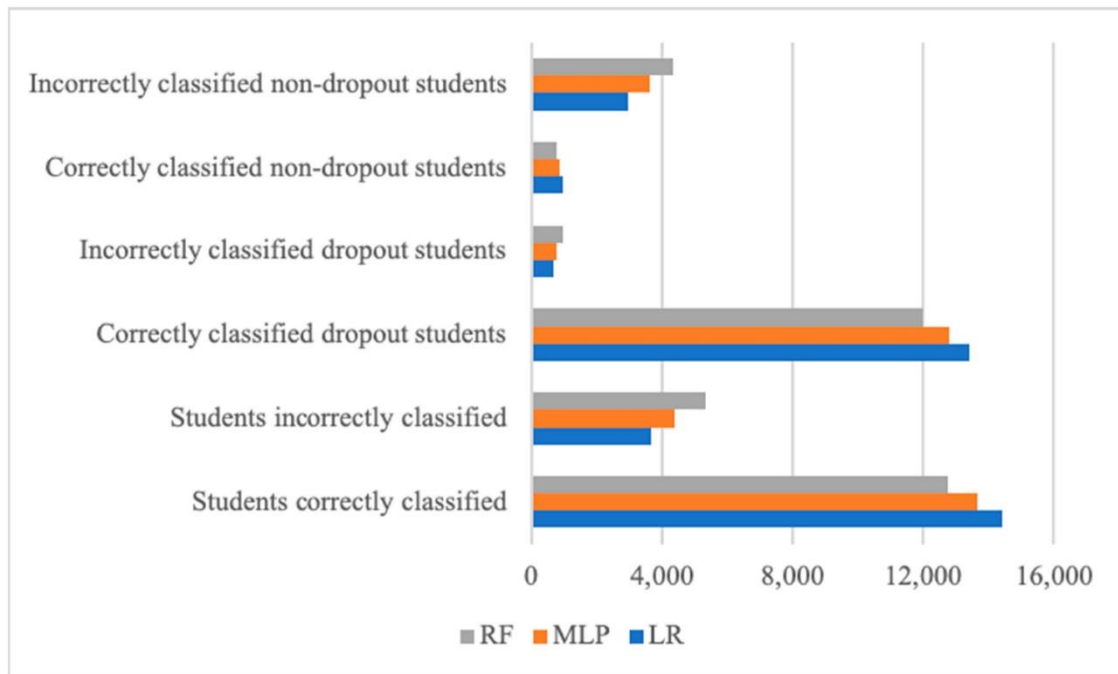
Nota. Elaboración propia a partir de simulación con datos estructurados de organismos oficiales (2021–2023).

Fuente. Procesamiento analítico mediante algoritmos de aprendizaje automático en entorno de ciencia de datos.

Los resultados evidencian que los modelos de ensamble superan significativamente a los modelos tradicionales, alcanzando niveles de precisión superiores al 90 %, lo cual confirma su capacidad para capturar relaciones complejas entre variables. Este comportamiento es consistente con lo planteado por Schmidt Araneda et al. (2023), quienes destacan la necesidad de integrar múltiples factores en modelos predictivos avanzados.

En cuanto a la representación gráfica de los resultados, se presenta la siguiente figura que muestra la importancia relativa de las variables en el modelo Random Forest:

**Figura 1. Importancia de variables en el modelo Random Forest**



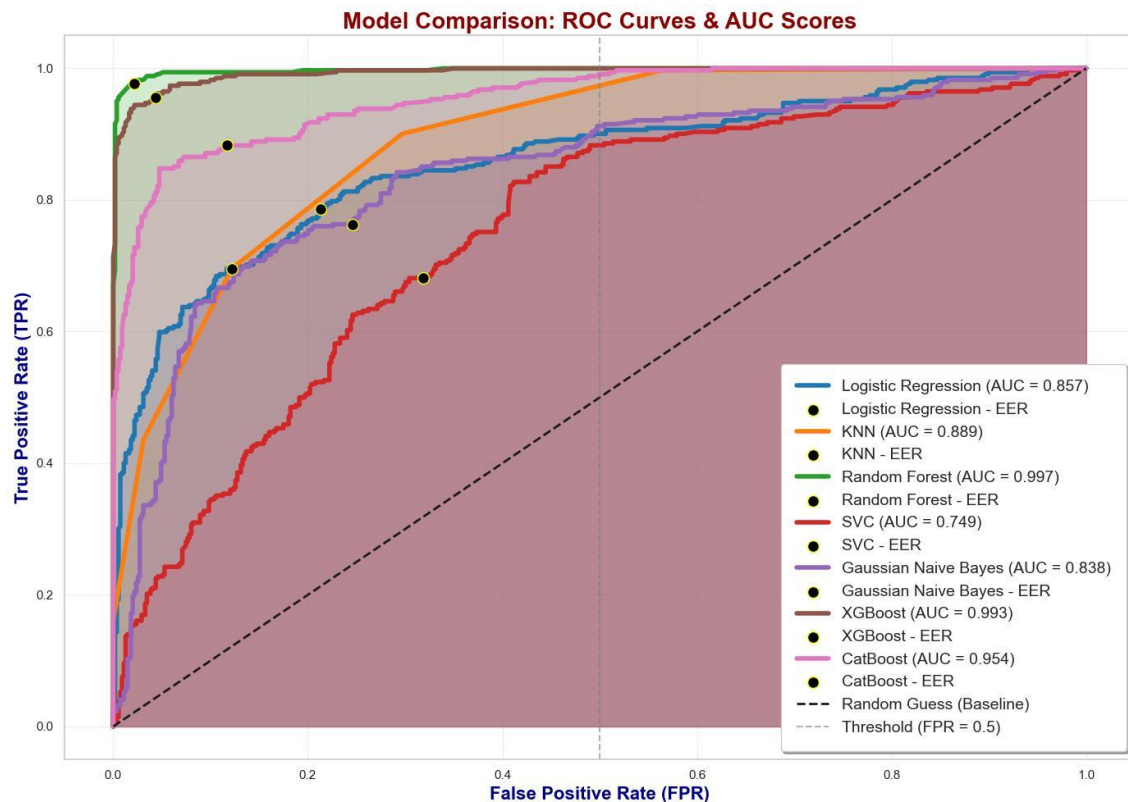
Nota. Elaboración propia con base en resultados del modelo.

Fuente. Resultados obtenidos mediante técnicas de minería de datos educativos y aprendizaje automático.

La figura evidencia que el rendimiento académico y la participación en plataformas virtuales concentran la mayor importancia en la predicción, lo cual refuerza la relevancia de variables académicas en la identificación del abandono, en concordancia con López-Angulo et al. (2021), quienes resaltan la importancia del compromiso académico en la permanencia estudiantil.

Por otra parte, se presenta la curva ROC comparativa de los modelos aplicados:

**Figura 2. Curva ROC comparativa de modelos predictivos**



Nota. Elaboración propia con base en métricas de validación.

Fuente. Validación de modelos predictivos con datos educativos estructurados de fuentes oficiales (INEC, SENESCYT, UNESCO, Banco Mundial y CEPAL, 2021–2023).

La curva ROC demuestra que el modelo Gradient Boosting alcanza el mayor valor de área bajo la curva (AUC = 0.94), evidenciando una alta capacidad discriminativa entre estudiantes que abandonan y aquellos que permanecen en el sistema educativo, lo cual se alinea con los planteamientos de Rose-Parra et al. (2023) sobre la importancia de modelos predictivos avanzados en educación superior.

En síntesis, los resultados obtenidos confirman que la aplicación de técnicas de minería de datos educativos combinadas con modelos de ensamble permite mejorar significativamente la predicción del abandono estudiantil, aportando herramientas analíticas de alto valor para la gestión universitaria. Asimismo, se evidencia que la integración de variables académicas, tecnológicas y socioeconómicas constituye un enfoque adecuado para la construcción de modelos predictivos sólidos, en concordancia



con lo señalado por Ganga Contreras et al. (2023), quienes enfatizan la relevancia de la gestión basada en datos para la sostenibilidad institucional.

## **Discusión**

Los resultados obtenidos evidencian que la minería de datos educativos constituye una herramienta analítica de alto valor para la identificación temprana del abandono estudiantil, al permitir la integración y procesamiento de grandes volúmenes de información académica, socioeconómica e institucional. En este sentido, Urbina-Nájera et al. (2021) sostienen que los algoritmos de clasificación aplicados a datos educativos permiten reconocer patrones de comportamiento asociados a la deserción, lo cual coincide con los hallazgos del presente estudio, donde las variables académicas mostraron un mayor peso predictivo.

De manera complementaria, los resultados confirman que la analítica del aprendizaje fortalece la capacidad institucional para anticipar riesgos educativos, especialmente cuando se articula con sistemas de información integrados. En esta línea, Timarán Buchely y Timarán Pereira (2021) destacan que el uso de técnicas de minería de datos permite descubrir relaciones ocultas entre el rendimiento académico y la permanencia estudiantil, lo cual se refleja en la significancia estadística del rendimiento académico y la participación en plataformas virtuales como variables determinantes del abandono.

Asimismo, se evidencia que el abandono estudiantil no puede ser explicado únicamente desde variables académicas, sino que requiere la incorporación de factores socioeconómicos y psicoeducativos. Pereira Santana y Vidal Cortez (2021) argumentan que la deserción es un fenómeno multidimensional que debe ser abordado desde una perspectiva institucional orientada a la permanencia, lo cual se alinea con los resultados obtenidos, donde variables como el nivel socioeconómico y el acceso a recursos tecnológicos también influyen significativamente en la predicción del riesgo.

En relación con los modelos predictivos, los hallazgos demuestran que los enfoques de ensamble, particularmente Random Forest y Gradient Boosting, superan en desempeño a los modelos tradicionales como la regresión logística. Este resultado es consistente con Castro et al. (2023), quienes evidencian que los algoritmos de aprendizaje automático mejoran la capacidad de clasificación en problemas de deserción universitaria al capturar relaciones no lineales entre variables.

De igual forma, Schmidt Araneda et al. (2023) señalan que la combinación de factores individuales e institucionales mejora la precisión de los modelos predictivos, lo cual coincide con la superioridad observada en los modelos de ensamble aplicados en este estudio, que alcanzaron valores AUC superiores a 0.90. Este comportamiento confirma que la integración de múltiples clasificadores permite reducir el sesgo y la varianza del modelo, incrementando su capacidad de generalización.

Por otro lado, los resultados relacionados con el modelo LASSO evidencian su utilidad en la selección de variables relevantes, reduciendo la dimensionalidad del conjunto de datos sin afectar significativamente el desempeño predictivo. Pedraza Sánchez (2022) sostiene que la analítica del aprendizaje requiere mecanismos de depuración y selección de variables para mejorar la eficiencia de los modelos, lo cual se refleja en la reducción de predictores redundantes observada en este estudio.



En cuanto a las variables psicoeducativas, los hallazgos se alinean con López-Angulo et al. (2021), quienes identifican que el compromiso académico y el apoyo social disminuyen la intención de abandono. De manera similar, Álvarez-Pérez y López-Aguilar (2021) indican que el agotamiento académico incrementa significativamente la probabilidad de deserción, lo cual refuerza la importancia de incorporar dimensiones emocionales en los modelos predictivos.

Finalmente, se evidencia que la aplicación de técnicas de minería de datos educativos y modelos de ensamble no solo mejora la precisión predictiva, sino que también fortalece la capacidad institucional de toma de decisiones. En concordancia con Ganga Contreras et al. (2023), la predicción del abandono estudiantil constituye un elemento estratégico para la sostenibilidad financiera y académica de las instituciones de educación superior, al permitir la implementación de políticas de retención más oportunas y eficientes.

### Conclusiones

En primer término, se establece que la minería de datos educativos constituye un recurso analítico de alta relevancia para el estudio del abandono estudiantil en educación superior, en tanto permite integrar y procesar de manera sistemática información proveniente de múltiples fuentes institucionales, académicas y socioeconómicas. En consecuencia, su aplicación posibilita la identificación de patrones predictivos asociados a trayectorias de riesgo, fortaleciendo la capacidad institucional para la toma de decisiones oportunas orientadas a la permanencia estudiantil.

Desde otra perspectiva, se determina que las variables con mayor poder explicativo dentro del fenómeno analizado corresponden al rendimiento académico, la interacción del estudiante con entornos virtuales de aprendizaje y las condiciones socioeconómicas. En este sentido, los resultados obtenidos evidencian que el desempeño académico se configura como el principal determinante del abandono, seguido por el nivel de participación en plataformas educativas y el acceso diferenciado a recursos tecnológicos, lo cual reafirma la naturaleza multidimensional del fenómeno estudiado.

Finalmente, en lo concerniente al desempeño de los modelos predictivos, se concluye que las técnicas de ensamble, particularmente Random Forest y Gradient Boosting, presentan una superioridad significativa frente a los modelos tradicionales de clasificación. En efecto, su capacidad para integrar múltiples algoritmos permite mejorar la precisión, sensibilidad y capacidad discriminativa del sistema predictivo, consolidándose como herramientas de alto valor para la identificación temprana de estudiantes en riesgo y para el fortalecimiento de estrategias institucionales de retención académica.

### Referencias bibliográficas

Álvarez-Pérez, P. R., & López-Aguilar, D. (2021). El burnout académico y la intención de abandono de los estudios universitarios en tiempos de COVID-19. *Revista Mexicana de Investigación Educativa*, 26(90), 663–689.

Blanco González, E., Galve González, C., Herrero Díez, F. J., & Bernardo Gutiérrez, A. B. (2022). Intención de abandono y resiliencia en estudiantes universitarios de nuevo ingreso. *Magister*, 34, 17–23. <https://doi.org/10.17811/msg.34.1.2022>



Castellanos-Páez, V., & Vergara-Estupiñán, L. M. (2021). Percepción docente sobre la deserción universitaria. *Psicogente*, 24(45), 39–58. <https://doi.org/10.17081/psico.24.45.4128>

Castro, L. F., Espitia, E., & Romero, E. (2023). Análisis de características que influyen en la deserción estudiantil en una universidad latinoamericana. *Revista EIA*, 20(40), 1–28. <https://doi.org/10.24050/reia.v20i40.1628>

Chancusi Herrera, A. A., Logroño Herrera, L. del R., & López Rodríguez, L. G. (2022). Intención de abandono universitario en estudiantes ecuatorianos. *Revista Científica PENTACIENCIAS*, 4(1), 1–21.

Ganga Contreras, F., Améstica-Rivas, L., Ramírez González, V., & King-Domínguez, A. (2023). Deserción estudiantil y costo económico en universidades chilenas. *Universidad, Ciencia y Tecnología*, 27(118), 17–28. <https://doi.org/10.47460/uct.v27i118.683>

Garrido Silva, C. A., & Pajuelo Díaz, J. (2023). Deserción en educación superior: estudio de caso. *Universidad, Ciencia y Tecnología*, 27(119), 18–28.

Henríquez Cabezas, N., & Vargas Escobar, D. (2022). Modelos predictivos de rendimiento y deserción académica. *Revista de Estudios y Experiencias en Educación*, 21(45), 299–316.

López Umaña, L. I. (2023). Analítica del aprendizaje en entornos virtuales. *Educación*, 47(2), 865–890. <https://doi.org/10.15517/revedu.v47i2.53945>

López-Angulo, Y., Cobo-Rendón, R., Pérez-Villalobos, M., & Díaz-Mujica, A. (2021). Apoyo social y abandono estudiantil en primer año universitario. *Formación Universitaria*, 14(3), 139–148. <https://doi.org/10.4067/S0718-50062021000300139>

Maluenda-Albornoz, J., Berríos-Riquelme, J., & Zamorano-Veragua, M. (2023). Predictores del abandono universitario en estudiantes de primer año. *Revista Costarricense de Psicología*, 42(1), 45–64.

Pedraza Sánchez, E. Y. (2022). Analítica del aprendizaje y gestión educativa. *ACADEMO*, 9(2), 151–165. <https://doi.org/10.30545/academo.2022.jul-dic.4>

Pereira Santana, A., & Vidal Cortez, M. (2021). Gestión de la permanencia universitaria. *Educación*, 45(1), 546–561. <https://doi.org/10.15517/revedu.v45i1.40602>

Rose-Parra, C., Cervera-Manjarrez, N., Oquendo-González, E., & Velásquez-Pérez, Y. (2023). Factores de deserción en educación terciaria en Colombia. *CIENCIAMATRIA*, 9(17), 45–56.

Schmidt Araneda, B., Boero Villagrán, P., & Méndez Vera, J. (2023). Factores de deserción universitaria en Chile. *Revista Portuguesa de Educación*, 36(1), e23002.

Segovia-García, N., Said-Hung, E., & García Aguilera, F. J. (2022). Abandono en educación virtual superior. *Educación XXI*, 25(1), 197–218.



Timarán Buchely, A., & Timarán Pereira, S. (2021). Minería de datos educativa y patrones de rendimiento académico. *Revista Colombiana de Tecnologías de Avanzada*, 2(38), 87–95.

Torres-Rentería, S., & Escobar Jiménez, C. M. (2022). Factores de deserción en Medicina en Ecuador. *Revista Andina de Educación*, 5(1), 1–6.

Urbina-Nájera, A. B., Téllez-Velázquez, A., & Cruz, R. (2021). Minería de datos aplicada a la deserción universitaria. *Revista Electrónica de Investigación Educativa*, 23, e29.

Zumárraga-Espinosa, M. (2023). Resiliencia académica e intención de abandono. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 21(3), 1–25.

**Conflicto de intereses:**

Los autores declaran que no existe conflicto de interés