

Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2024,
Volumen 8, Número 6.

https://doi.org/10.37811/cl_rcm.v8i6

**EL MACHINE LEARNING PARA ABORDAR EL
ABANDONO ESCOLAR: UNA REVISIÓN DE LOS
MODELOS MÁS INNOVADORES**

**MACHINE LEARNING TO ADDRESS SCHOOL DROPOUT: A
REVIEW OF THE MOST INNOVATIVE MODELS**

Jules Mao Flores Satalaya

Universidad Nacional Mayor de San Marcos, Lima, Perú

El machine learning para abordar el abandono escolar: Una revisión de los modelos más innovadores

Jules Mao Flores Satalaya

jules.flores@unmsm.edu.pe

<https://orcid.org/0009-0001-6553-0809>

Universidad Nacional Mayor de San Marcos, Lima, Perú

RESUMEN

El objetivo de este estudio es analizar el papel del machine learning como herramienta innovadora para identificar, predecir y abordar el abandono escolar, evaluando los modelos más efectivos y su aplicación en contextos educativos. Para ello, se llevó a cabo una revisión sistemática en las bases de datos Scopus, Web of Science (WOS) y SciELO, utilizando la metodología PRISMA y adoptando un enfoque cualitativo y descriptivo. Los criterios de inclusión comprendieron trabajos publicados entre 2020 y 2024, estudios originales, artículos en inglés o español y documentos con texto completo que estuvieran directamente relacionados con el objetivo del estudio. Se emplearon palabras clave como "machine learning", "machine learning algorithms", "learning algorithms", "machine learning models", "learning models", "school dropout", "school abandonment", "student dropout", "student attrition". De los 773 documentos identificados, se eliminaron 753 por no cumplir con los criterios establecidos, resultando en 20 artículos seleccionados para su análisis. Los resultados muestran que modelos como las Redes Neuronales Artificiales (ANN), Vecinos Más Cercanos (KNN), Regresión Lineal (LR) y Árboles de Decisión (DT) han demostrado eficacia en la clasificación y predicción. El rendimiento académico previo es un predictor clave del abandono escolar, junto con factores como dificultades financieras y la falta de apoyo social, que afectan la permanencia de los estudiantes. En conclusión, el machine learning (ML) en la educación resalta su capacidad para identificar y prevenir el abandono escolar. Las técnicas de ML permiten a las instituciones predecir con precisión los riesgos de deserción y desarrollar intervenciones personalizadas para los estudiantes.

Palabras clave: modelos, machine learning, abandono, deserción, educación

Machine learning to address school dropout: A review of the most innovative models

ABSTRACT

The objective of this study is to analyze the role of machine learning as an innovative tool to identify, predict, and address school dropout, evaluating the most effective models and their application in educational contexts. To achieve this, a systematic review was conducted in the Scopus, Web of Science (WOS), and SciELO databases, using the PRISMA methodology and adopting a qualitative and descriptive approach. The inclusion criteria comprised works published between 2020 and 2024, original studies, articles in English or Spanish, and documents with full text that were directly related to the study's objective. Keywords such as "machine learning," "machine learning algorithms," "learning algorithms," "machine learning models," "learning models," "school dropout," "school abandonment," "student dropout," and "student attrition" were employed. Of the 773 documents identified, 753 were eliminated for not meeting the established criteria, resulting in 20 articles selected for analysis. The results show that models such as Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Linear Regression (LR), and Decision Trees (DT) have proven effective in classification and prediction. Prior academic performance is a key predictor of school dropout, along with factors such as financial difficulties and lack of social support, which affect students' retention. In conclusion, machine learning (ML) in education highlights its capacity to identify and prevent school dropout. ML techniques enable institutions to accurately predict dropout risks and develop personalized interventions for students.

Keywords: models, machine learning, abandonment, dropout, education

Artículo recibido 02 diciembre 2024

Aceptado para publicación: 28 diciembre 2024



INTRODUCCIÓN

La inteligencia artificial (IA) es una parte amplia de la ciencia de la computación que se ocupa de construir máquinas inteligentes equipadas para realizar tareas que normalmente requieren inteligencia humana. Por su parte, los sistemas de IA funcionan al ingerir grandes cantidades de datos de entrenamiento etiquetados, analizando los datos en busca de patrones y correlaciones, y utilizando estos patrones para hacer pronósticos sobre estados futuros (Nimbalkar & Berad, 2021). El concepto de inclusión de la IA y el aprendizaje automático (ML) es extremadamente poderoso: ambos permiten que los sistemas se programen a sí mismos y mejoren su rendimiento a través de un proceso de mejora continua (Sahana et al., 2023).

En el ámbito educativo, el machine learning (ML) ha adquirido una relevancia considerable. La incorporación de la tecnología en la educación es cada vez más común, y ya se están implementando numerosas soluciones, al tiempo que se exploran muchas otras. De este modo, el aprendizaje automático ha tenido un impacto significativo en el sector educativo (Alhabeeb et al., 2024; Okagbue et al., 2023; Tiwari, 2023). Dentro de la educación, el aprendizaje automático se ha utilizado para modelar sistemas de tutoría inteligentes para estudiantes, hacer recomendaciones sobre posibles trayectorias profesionales futuras, mejorar el plan de estudios, predecir con precisión las calificaciones de los estudiantes y mucho más (Razaulla et al., 2022; Wu, 2020).

Durante muchos años, el aprendizaje automático, ha seguido siendo un campo de desarrollo vital, ya que es la medida clave de la civilización y la evolución del mundo, con su enorme efecto tanto en los individuos como en las sociedades (Alenezi & Faisal, 2020; Shah et al., 2021). Con el rápido desarrollo de la tecnología de la información y la inteligencia artificial, el campo de la educación avanza gradualmente hacia la educación inteligente (Zhou & Song, 2020). El machine learning y otros campos relacionados con la ciencia de datos y la inteligencia artificial, permiten el desarrollo de técnicas avanzadas que aportan un gran potencial, dando lugar a nuevas aplicaciones y aproximaciones más efectivas en el análisis y la predicción académica (Gómez-Pulido et al., 2023).

Por otro lado, la deserción estudiantil representa un problema grave a nivel mundial, con profundas implicaciones para los estudiantes, las instituciones educativas y la sociedad en su conjunto (Hassan et al., 2024; Matz et al., 2023). Este fenómeno es un gran desafío, ya que aproximadamente el 40% de los estudiantes en EE. UU. no logran completar sus estudios (Vaarma & Li, 2024; Villegas-Ch et al., 2023).



Asimismo, en Somalilandia, las altas tasas de deserción escolar constituyen un problema crítico (Hassan et al., 2024). De manera análoga, en Egipto, el abandono escolar afecta gravemente a todos los sistemas educativos del país (Selim & Rezk, 2023). En Europa, cada año, miles de jóvenes abandonan la escuela sin haber completado la educación que iniciaron; esta situación resulta en la falta de oportunidades de formación, perspectivas laborales limitadas y, a menudo, conduce al estancamiento profesional y a la exclusión social (European Commission, 2024).

Por consiguiente, la predicción automatizada de la retención y graduación estudiantil mediante Machine Learning ha ganado un creciente interés entre educadores tanto en teoría como en práctica (Okoye et al., 2024a). A pesar de que el abandono escolar representa un desafío complejo en el sector educativo, en parte debido al desequilibrio en los datos relacionados con este fenómeno (Mduma, 2023), los modelos de aprendizaje automático (ML) han demostrado ser eficaces para identificar a los estudiantes en riesgo de abandonar sus estudios con un alto grado de precisión (Fauszt et al., 2023). Predecir el comportamiento estudiantil es, por lo tanto, una tarea crucial para las instituciones educativas (Villar & Velini, 2024). Debido a la complejidad de la toma de decisiones humanas, basarse en interpretaciones a través de modelos de aprendizaje automático puede generar ventajas competitivas en el ámbito educativo (Delen et al., 2023). La revisión del machine learning como herramienta para abordar el abandono escolar ofrece una perspectiva integral sobre cómo las técnicas avanzadas de análisis de datos pueden ser aplicadas para identificar y mitigar este fenómeno crítico. El abandono escolar es un problema multidimensional que afecta a estudiantes en diversas etapas educativas, impactando no solo su futuro académico y profesional, sino también el desarrollo socioeconómico de las comunidades. En este contexto, es fundamental examinar la relevancia del machine learning, que ha demostrado su eficacia en la predicción de comportamientos y la personalización de intervenciones. Al revisar los modelos más innovadores, se busca aclarar conceptos clave, evaluar avances recientes y proporcionar estrategias basadas en evidencia que puedan ayudar a las instituciones educativas a desarrollar programas más efectivos para retener a los estudiantes y fomentar su éxito académico. Esta revisión no solo aborda la necesidad de soluciones prácticas, sino que también contribuye a un entendimiento más profundo de cómo la tecnología puede influir en la educación y el bienestar de los estudiantes.

En tal sentido el artículo, se plantea como objetivo: Analizar el papel del machine learning como herramienta innovadora para identificar, predecir y abordar el abandono escolar, evaluando los modelos más efectivos y su aplicación en contextos educativos. Asimismo, como objetivos específicos: Identificar los modelos de machine learning más efectivos en la predicción del abandono escolar y sus resultados en la investigación. Examinar los factores que contribuyen al abandono escolar y cómo se integran en los modelos de machine learning. Evaluar cómo la implementación de machine learning puede ayudar a las instituciones educativas a desarrollar estrategias de intervención para reducir el abandono escolar.

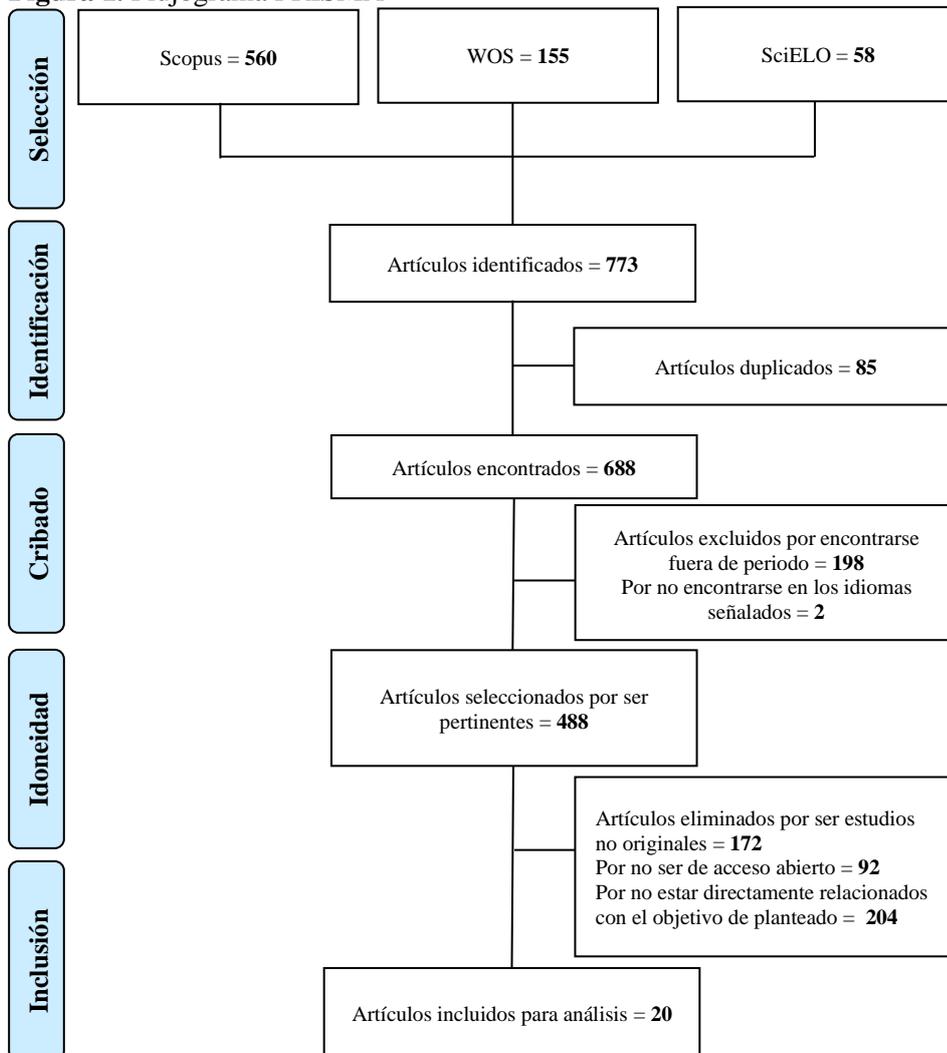
MATERIALES Y MÉTODOS

Una revisión sistemática es una síntesis (combinación, amalgama, fusión) de la literatura de investigación, cuyo objetivo principal es responder de manera precisa a una pregunta específica. Este tipo de revisión sigue un proceso metódico y estructurado que permite identificar, seleccionar, evaluar y sintetizar toda la evidencia disponible de alta calidad relevante para la cuestión planteada (Bettany-Saltikov & McSherry, 2024). Para ello, se realizan búsquedas exhaustivas que permiten identificar todos los estudios que cumplan con criterios de elegibilidad previamente establecidos (Sinha et al., 2022). En la revisión sistemática presentada en este artículo, se analizaron y sintetizaron investigaciones relacionadas con el uso del machine learning como herramienta para enfrentar el problema del abandono escolar, consolidando las evidencias encontradas en diferentes estudios.

Se realizó la búsqueda en base de datos distinguidas como SCOPUS, Web of Science (WOS) y Scielo, mediante operadores booleanos AND y OR, equivalente al siguiente query: (("machine learning") OR ("machine-learning") OR ("aprendizaje automático") OR ("deep learning") OR ("ML") OR ("machine learning algorithms") OR ("learning algorithms") OR ("machine learning models") OR ("learning models") OR ("algoritmos de machine learning") OR ("modelos de machine learning")) AND (("school dropout") OR ("school abandonment") OR ("student dropout") OR ("student attrition") OR ("school leaving") OR ("student retention") OR ("deserción") OR ("abandono") OR ("deserción educativa") OR ("deserción escolar") OR ("abandono educativo") OR ("abandono escolar") OR ("student income") OR ("educational income") OR ("school rent") OR ("retención estudiantil") OR ("retención educativa") OR ("retención escolar") OR ("retención")). Se consideraron publicaciones desde el año 2020 hasta el 2024, todos de acceso

abierto (Open Access). Para la inclusión y exclusión de artículos fue de gran apoyo la metodología PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyse) (Figura 1).

Figura 1. Flujograma PRISMA



En el proceso de selección de artículos utilizando la metodología PRISMA, se consultaron tres bases de datos: Scopus, Web of Science (WOS) y SciELO, identificando inicialmente 773 artículos (560 de Scopus, 155 de WOS y 58 de SciELO). Tras eliminar 85 artículos duplicados, quedaron 688 documentos para el cribado, donde se excluyeron 198 artículos por estar fuera del período estipulado y 2 más por no estar en los idiomas señalados, reduciendo el número a 488. En la fase de evaluación de idoneidad, se eliminaron 172 artículos por no ser estudios originales, 92 por no ser de acceso abierto y 203 por no estar directamente relacionados con el objetivo de la investigación. Finalmente, se incluyeron 20 artículos que cumplían con los criterios para ser analizados en el estudio.

RESULTADOS

La revisión sistemática realizada reveló un patrón interesante en la distribución de investigaciones sobre modelos de machine learning para abordar el abandono escolar. Se observó que Scopus concentra la mayor parte de los estudios, representando un significativo 90% del total de investigaciones incluidas en el análisis.

Esto destaca la relevancia de Scopus como una fuente principal de literatura científica en este ámbito.

Los resultados obtenidos en la Tabla 1, revelaron una clara concentración de investigaciones en la base de datos Scopus, que alberga el 90% (18) de los estudios analizados, lo que subraya su preeminencia como fuente de información científica en este ámbito. En contraste, la base de datos Web of Science (WOS) cuenta con apenas el 10% (2) de los estudios, lo que sugiere una menor representación en este campo de investigación específica. Finalmente, en SciELO no se encontraron estudios que cumplieran con los criterios de inclusión establecidos para el análisis, lo que indica una posible brecha en la disponibilidad de investigaciones relacionadas en esta plataforma. Estos hallazgos destacan la disparidad en la distribución de investigaciones en las diferentes bases de datos, lo que puede tener implicaciones para la accesibilidad y visibilidad de la literatura científica sobre los modelos de algoritmos de machine learning para abordar el abandono escolar.

Tabla 1 Artículos según base de datos

Base de datos	Cantidad artículos	Incluidos	%
Scopus	560	18	90%
Wos	155	2	10%
SciELO	58	0	0%
Total	773	20	100%

La revisión de la disponibilidad de estudios según el año de publicación muestra un notable incremento en el uso de modelos de machine learning para abordar el problema del abandono escolar. Según la Tabla 2, en 2020 solo se identificó un estudio en las bases de datos seleccionadas, representando un 5% del total, lo que indica un interés inicial en el tema. En 2021, la producción aumentó a 3 estudios (15%), reflejando un interés creciente. Este incremento continuó en 2022 con 3 publicaciones adicionales (15%), consolidando la atención en esta línea de investigación.

El 2023 marcó un cambio significativo, con 9 publicaciones (45%), el mayor número registrado hasta ahora. Aunque el año 2024 aún no ha concluido, ya se cuenta con 4 estudios (20%), lo que eleva el total de investigaciones revisadas a 20. Este incremento sostenido destaca el creciente interés y compromiso de la comunidad científica en el uso de machine learning para abordar el abandono escolar, subrayando la relevancia y potencial impacto de estos modelos en la actualidad.

Tabla 2 Artículos según año de publicación

Base de datos	Incluidos	%
2024	4	20%
2023	9	45%
2022	3	15%
2021	3	15%
2020	1	5%
Total	20	100.0

En la Tabla 3 se presenta la distribución de artículos según el país de estudio. De un total de 20 artículos, México destaca como el país con mayor representación, con 3 artículos (15%). Le siguen Perú, Estados Unidos, Brasil, España, Tanzania, Eslovaquia y Chile, cada uno con 2 artículos (10%). En menor proporción, Finlandia, Ecuador y Hungría contribuyen con 1 artículo cada uno (5%). Esta distribución muestra a México como la principal fuente de los artículos revisados, mientras que los demás países tienen una participación similar, aunque más limitada.

Tabla 3 Artículos según país de estudio

País	Cantidad artículos	%
México	3	15%
Perú	2	10%
EEUU	2	10%
Brasil	2	10%
España	2	10%
Tanzania	2	10%
Eslovaquia	2	10%
Chile	2	10%

Finlandia	1	5%
Ecuador	1	5%
Hungría	1	5%
Total	20	100.0 %

Asimismo, para registrar los datos se utilizó un protocolo que permitió organizar la información de cada artículo. El protocolo recoge información de los siguientes campos: Título, Autor(es), Revista, País, Año y Conclusión (Tabla 4).

Tabla 4 Artículos científicos incluidos en la revisión según las palabras clave utilizadas

Título	Autor(es)	Revista	País	Año	Conclusión
Application of the performance of machine learning techniques as support in the prediction of school dropout	Jiménez-Gutiérrez, Auria Lucia Mota-Hernández, Cinthya Ivonne Mezura-Montes, Efrén Alvarado-Corona, Rafael	Scientific Reports	México	2024	Para abordar la deserción en niveles educativos medio y superior en México, se implementaron técnicas de Machine Learning que lograron alta precisión. Las redes neuronales artificiales (ANN) alcanzaron un 99% de confiabilidad, mientras que el uso de máquinas de soporte vectorial, optimización bayesiana, Random Forest y regresiones Ridge y Lasso lograron una confiabilidad superior al 91%.
Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education	Okoye, Kingsley Nganji, Julius T. Escamilla, Jose Hosseini, Samira	Computers and Education: Artificial Intelligence	México	2024	Se desarrolló un modelo de Machine Learning (RG-DMML) y un algoritmo de conjunto para predecir la retención y graduación de estudiantes en educación, basado en la metodología CRISP-DM. Utilizando el método Bagging, selección de características Wrapper y

Predicting and Mitigating Freshmen Student Attrition: A Local-Explainable Machine Learning Framework	Delen, Dursun Davazdahemami, Behrooz Rasouli Dezfouli, Elham	Information Systems Frontiers	EEUU	2024	<p>el algoritmo K-Nearest Neighbor (KNN) con validación cruzada k-fold, el modelo demostró ser eficiente y preciso en la predicción de estos resultados estudiantiles.</p> <p>Usando un enfoque de red neuronal profunda con perceptrón multicapa (MLP) completamente conectado (o denso) para desarrollar un modelo capaz de predecir la retención estudiantil, considerando factores personales, educativos, financieros y socioeconómicos.</p>
Predicting student dropouts with machine learning: An empirical study in Finnish higher education	Vaarma, Matti Li, Hongxiu	Technology in Society	Finlandia	2024	<p>Utilizando tres modelos de machine learning: CatBoost (CAT), redes neuronales (NN) y regresión logística (LR), se obtuvo información sobre la confiabilidad con la que es posible predecir el abandono escolar en un momento determinado mediante estos algoritmos.</p>
An explainable machine learning approach for student dropout prediction	Krüger, João Gabriel Corrêa Britto, Alceu de Souza Barddal, Jean Paul	Expert Systems with Applications	Brasil	2023	<p>Los algoritmos Random Forest, XGBoost, Gradient Boosting y AdaBoost, obtuvieron puntuaciones AUC-PR entre el 38.22 % y el 89.50 % al predecir la deserción en distintos momentos del año.</p>

<p>A Case-Study Comparison of Machine Learning Approaches for Predicting Student's Dropout from Multiple Online Educational Entities</p>	<p>Porras, Manuel Lara, Juan Alfonso Romero, Cristóbal Ventura, Sebastián</p>	<p>Algorithms</p>	<p>España</p>	<p>2023</p>	<p>Se logró predecir la deserción estudiantil en cursos en línea mediante el uso de una red neuronal de memoria a largo y corto plazo bidireccional (BiLSTM).</p>
<p>Using machine learning to predict student retention from socio- demographic characteristics and app-based engagement metrics</p>	<p>Matz, Sandra C. Bukow, Christina S. Peters, Heinrich Deacons, Christine Dinu, Alice Stachl, Clemens</p>	<p>Scientific Reports</p>	<p>EEUU</p>	<p>2023</p>	<p>A través de dos algoritmos, Elastic Net y Random Forest, se obtuvieron predicciones confiables. Destacamos la capacidad de generalización de resultados, ya que los modelos entrenados en una universidad pueden predecir la retención en otra institución con niveles de desempeño predictivo razonablemente altos.</p>
<p>Improving Student Retention in Institutions of Higher Education through Machine Learning: A Sustainable Approach</p>	<p>Villegas-Ch, William Govea, Jaime Revelo-Tapia, Solange</p>	<p>Sustainability</p>	<p>Ecuador</p>	<p>2023</p>	<p>La investigación confirma que la implementación del modelo de predicción de retención estudiantil basado en machine learning mediante ANN (Red Neuronal Artificial) ha tenido un impacto significativo, con un aumento del 6.875% en la tasa de retención. Esto resalta la efectividad del</p>

Data Balancing Techniques for Predicting Student Dropout Using Machine Learning Comparative analysis of Machine Learning Techniques for the prediction of cases of university dropout	Mduma, Neema	Data	Tanzania	2023	<p>modelo para identificar de manera temprana a los estudiantes en riesgo y mejorar así la retención.</p> <p>El método SMOTE con vecinos más cercanos (KNN) editados logró el mejor rendimiento de clasificación en la muestra de retención en 10 ocasiones.</p>
Design of a Machine Learning Model to Predict Student Attrition	Fauszt, Tibor Erdélyi, Katalin Dobák, Dóra Bognár, László Kovács, Endre	RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao	Perú	2023	<p>Se concluyó que la regresión logística es la técnica que ofrece los mejores resultados para predecir la deserción universitaria en el conjunto de datos analizado.</p> <p>El modelo de árbol de decisión desarrollado a partir de los datos de entrenamiento mostró una precisión de clasificación del 87,9% al ser evaluado con los datos de prueba. Este modelo fue diseñado específicamente para predecir situaciones en las que se anticipa que un estudiante abandonará la escuela y no completará su graduación.</p>
Classification model for student	Villarreal-Torres, Henry	EAI Endorsed Transactions	Perú	2023	<p>La investigación dio como resultado un modelo de aprendizaje automático</p>

dropouts using machine learning: case study	Ángeles-Morales, Julio Cano-Mejía, Jenny Mejía-Murillo, Carmen Flores-Reyes, Gumercindo Palomino-Márquez, Manuel Marín-Rodríguez, William Andrade-Girón, Daniel	on Scalable Information Systems			supervisado, conocido como Gradient Boosting Machine (GBM), para clasificar la deserción estudiantil. Este modelo identificó los principales factores asociados que influyen en la deserción, logrando un coeficiente de Gini del 92,20%, un AUC del 96,10% y un LogLoss de 24,24%. Estos resultados indican que el modelo tiene un desempeño eficiente y robusto.
Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role?	Segura, Marina Mello, Jorge Hernández, Adolfo	Mathematics	España	2022	En términos generales, todos los métodos de Machine Learning generan predicciones significativas. No obstante, hay algunas excepciones notables; por ejemplo, los métodos KNN muestran un rendimiento destacado dentro del contexto estudiado.
Enhanced Model for Predicting Student Dropouts in Developing Countries Using Automated Machine	Mnyawami, Yuda N. Maziku, Hellen H. Mushi, Joseph C.	Applied Artificial Intelligence	Tanzania	2022	El modelo propuesto alcanzó una precisión de predicción notable, con resultados de Árbol de Decisión (DT) = 99,8 %, K-Nearest Neighbors (KNN) = 99,6 %, Perceptrón Multicapa (MLP) = 99 % y Naive Bayes (NB) = 97 %. Esta mejora en la puntuación

Learning Approach: A Case of Tanzanian's Secondary Schools					de predicción sugiere una selección efectiva de las características que contribuyen al abandono escolar, las cuales pueden ser monitoreadas de cerca durante el proceso de aprendizaje para facilitar intervenciones tempranas
Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization Analysis of first-year university student dropout through machine learning models: A comparison between universities Towards predicting student's dropout in university	Niyogisubizo, Jovial Liao, Lyuchao Nziyumva, Eric Murwanashyaka, Evariste Nshimyumukiza, Pierre Claver Opazo, Diego Moreno, Sebastián Álvarez-Miranda, Eduardo Pereira, Jordi Kabathova, Janka Drlik, Martin	Computers and Education: Artificial Intelligence Mathematics	Eslovaquia	2022	A través de un modelo híbrido que combina Random Forest (RF), XGBoost, Gradient Boosting (GB) y Redes Neuronales Feedforward (FNN), es posible identificar a los estudiantes en riesgo de abandonar la escuela, considerando diversos factores influyentes. Entre los ocho modelos de aprendizaje automático evaluados en los conjuntos de datos, el algoritmo Gradient Boosting Tree demostraron ser el modelo más eficaz en la predicción de la deserción estudiantil universitaria. Los algoritmos de aprendizaje automático, como la Regresión Logística (LR), el Árbol de Decisión (DT), el Bosque Aleatorio

<p>courses using different machine learning techniques</p>		<p>(RF), el Bayes Ingenuo (NB), las Máquinas de Vectores de Soporte (SVM) y las Redes Neuronales (NN), demostraron que pueden aplicarse con éxito a un conjunto de datos, para predecir la deserción.</p>
<p>IoT system for school dropout prediction using machine learning techniques based on socioeconomic data</p>	<p>A.da S. Vasconcelos, Francisco F.X. Peixoto, Solon A. Hassan, Mohammad Mehedi Ali Akber Dewan, M.</p> <p>Electronics Brasil 2020</p>	<p>En el estudio se emplearon diversos clasificadores dentro de la deserción estudiantil, cada uno con características específicas: Árbol de Decisión, Regresión Logística (LR), Máquinas de Vectores de Soporte (SVM), K-Vecinos Más Cercanos (KNN), Perceptrón Multicapa (MLP) y Redes Neuronales Profundas (DNN). El Árbol de Decisión logró una precisión del 99,34% y una puntuación F1 del 100%, junto con un 98,69% de precisión en la recuperación.</p>
<p>Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and</p>	<p>Palacios, Carlos A. Reyes-Suárez, José A. Bearzotti, Lorena A. Leiva, Víctor Marchant, Carolina</p> <p>Entropy Chile 2021</p>	<p>Entre los algoritmos de aprendizaje automático empleados en este estudio se incluyen los árboles de decisión, k-vínculos más cercanos, regresión logística, Bayes ingenuo, bosque aleatorio y máquinas de vectores de soporte. De todos ellos, la técnica de bosque aleatorio demostró</p>

case study in chile					ser la más eficaz en términos de rendimiento para la predicción.
Ciencia de Datos Educativos y aprendizaje automático: un caso de estudio sobre la deserción estudiantil universitaria en México	Kuz, Antonieta Morales, Rosa	Education in the Knowledge Society	Mexico	2023	En el estudio se probaron técnicas como XGBoost, regresión logística, red neuronal y árboles de decisión. La red neuronal destaca por su área bajo la curva (AUC), que muestra una mayor distancia respecto a la línea de referencia, lo cual indica un rendimiento superior de este modelo en comparación con los otros para predecir la deserción estudiantil en dicho contexto.

La Tabla 5 presenta la distribución de estudios según el nivel educativo abordado, mostrando una mayor concentración en el nivel superior, con 16 estudios en total, entre los que destacan los trabajos de Okoye et al. (2024), Delen et al. (2023), Vaarma y Li (2024), entre otros, lo que refleja un interés significativo en la investigación de este nivel. En el nivel secundario, se registraron dos estudios, realizados por Krüger et al. (2023) y Mnyawami et al. (2022), indicando una menor atención en comparación con el nivel superior. Por su parte, el nivel multinivel fue abordado en un estudio de Jiménez-Gutiérrez et al. (2024), lo cual permite una visión más amplia que abarca varias etapas educativas. Finalmente, el nivel primario contó con un único estudio realizado por Mduma (2023), evidenciando una representación limitada en este segmento. En conjunto, estos resultados destacan un mayor enfoque investigativo en el nivel superior, mientras que los otros niveles muestran una menor presencia en el conjunto de estudios analizados.

Tabla 5 Estudios según nivel educativo

Nivel	Estudios
Superior	(Okoye et al., 2024); (Delen et al., 2023); (Vaarma & Li, 2024); (Porras et al., 2023); (Matz et al., 2023); (Villegas-Ch et al., 2023); (Tito et al., 2023); (Fauszt et al., 2023);

	(Villarreal-Torres et al., 2023); (Segura et al., 2022); (Niyogisubizo et al., 2022);
	(Opazo et al., 2021); (Kabathova & Drlik, 2021); (Freitas et al., 2020); (Palacios et al., 2021);
	(Kuz & Morales, 2023)
Secundaria	(Krüger et al., 2023); (Mnyawami et al., 2022);
Multinivel	(Jiménez-Gutiérrez et al., 2024);
Primaria	(Mduma, 2023)

La Tabla 5 presenta las metodologías de análisis de datos utilizadas en los estudios revisados, destacando que la metodología empírica es la más común, con 16 estudios aplicándola, lo que refleja una preferencia por enfoques basados en datos reales. En contraste, las metodologías CRISP-DM y KDD son menos frecuentes, cada una utilizada en solo dos estudios. Esto sugiere que, aunque CRISP-DM y KDD ofrecen enfoques estructurados para el análisis de datos, la mayoría de los estudios se centran en la recopilación y análisis empírico, lo que podría indicar áreas de mejora en la adopción de estas metodologías más formales en investigaciones futuras.

Tabla 6 Metodología de análisis de datos

Metodología de análisis de datos	Estudios
Empírica	(Delen et al., 2023); (Vaarma & Li, 2024); (Porrás et al., 2023); (Matz et al., 2023); (Villegas-Ch et al., 2023); (Tito et al., 2023); (Fauszt et al., 2023); (Villarreal-Torres et al., 2023); (Segura et al., 2022); (Niyogisubizo et al., 2022); (Opazo et al., 2021); (Freitas et al., 2020); (Kuz & Morales, 2023); (Krüger et al., 2023); (Mnyawami et al., 2022); (Mduma, 2023)
CRISP-DM	(Okoye et al., 2024); (Kabathova & Drlik, 2021)
KDD	(Jiménez-Gutiérrez et al., 2024); (Palacios et al., 2021)

La Tabla 7 presenta los modelos de machine learning más efectivos para predecir el abandono escolar, con un total de diez estudios analizados. En este contexto, los Árboles de Decisión (DT) sobresalen por su alta prevalencia, lo que resalta su popularidad, robustez y eficacia en la predicción. A continuación, se encuentran las Redes Neuronales Artificiales (ANN), respaldadas por siete estudios, lo que indica también

su notable eficacia. Por otro lado, la Regresión Lineal (LR) es citada en cuatro estudios, mientras que los Vecinos Más Cercanos (KNN) aparecen en tres estudios; esto sugiere que, aunque ambos son modelos válidos, su rendimiento es inferior al de los anteriores. La Máquina de Soporte de Vectores (SVM) y las Redes Bayesianas son mencionadas en un solo estudio cada una, lo que indica un uso más limitado en esta área. Por último, se menciona un modelo híbrido en un estudio, lo que sugiere que se están investigando combinaciones de diferentes enfoques de machine learning para mejorar la predicción del abandono escolar.

Tabla 7 Modelos de machine learning más efectivos en la predicción del abandono escolar

Modelos	Estudios
Redes Neuronales Artificiales (ANN)	(Jiménez-Gutiérrez et al., 2024); (Delen et al., 2023); (Vaarma & Li, 2024); (Porras et al., 2023); (Villegas-Ch et al., 2023); (Kabathova & Drlik, 2021); (Kuz & Morales, 2023)
Vecinos Mas Cercanos (KNN)	(Okoye et al., 2024); (Mduma, 2023); (Segura et al., 2022)
Regresión lineal (LR)	(Vaarma & Li, 2024); (Matz et al., 2023); (Tito et al., 2023); (Kabathova & Drlik, 2021);
Arboles de decisión (DT)	(Vaarma & Li, 2024); (Krüger et al., 2023); (Matz et al., 2023); (Fauszt et al., 2023); (Villarreal-Torres et al., 2023); (Mnyawami et al., 2022); (Opazo et al., 2021); (Kabathova & Drlik, 2021); (Freitas et al., 2020); (Palacios et al., 2021);
Máquina de Soporte de Vectores (SVM)	(Kabathova & Drlik, 2021)
Redes Bayesianas	(Kabathova & Drlik, 2021)
Hibrido	(Niyogisubizo et al., 2022)

En la Tabla 8 se presentan los algoritmos de machine learning más populares utilizados en la predicción del abandono escolar, destacándose la Red Neuronal Artificial (ANN) y el Árbol de Decisión, ambos con una frecuencia de 9, lo que sugiere su amplia adopción y efectividad en este ámbito. Le siguen la Regresión Logística (LR) y el Random Forest (RF), cada uno con 8, mostrando también un sólido desempeño en las predicciones. Otros algoritmos como la Máquina de Soporte de Vectores (SVM) y los Vecinos más Cercanos (KNN) tienen una frecuencia de 7, evidenciando su relevancia. A pesar de que métodos como XGBoost y Gradient Boosting están presentes, su uso es menos frecuente, con 3 y 4 respectivamente,

mientras que algoritmos como CatBoost y Elastic Net son poco utilizados, indicando que los enfoques más tradicionales siguen siendo preferidos en el análisis de abandono escolar.

Tabla 8 Algoritmos de machine learning más populares en la predicción del abandono escolar

Algoritmos	Frecuencia
Red Neuronal Artificial (ANN)	9
Máquina de Soporte de Vectores (SVM)	7
Vecinos más Cercanos (KNN)	7
Regresión Logística (LR)	8
Random Forest (RF)	8
Naive Bayes (NB)	6
Perceptrón Multicapa (MP)	4
Árbol de Decisión (DT)	9
Gradient Boosting (GB)	4
XGBoost	3
CatBoost (CAT)	1
Elastic Net	1
Aprendizaje Profundo	1
Regresión Lineal Ridge y Lasso	1
AdaBoost	1

En la Tabla 7, diversos estudios destacan que los factores más relevantes y repetidos en la deserción escolar incluyen principalmente elementos académicos, económicos, psicosociales, personales, demográficos e institucionales. En términos académicos, las calificaciones bajas, los cursos reprobados y los créditos acumulados se identifican como indicadores clave de riesgo, mientras que los problemas financieros, ingresos familiares bajos y la situación económica general son barreras significativas que afectan la permanencia escolar. Además, aspectos psicosociales, como las dificultades emocionales, la baja autoestima y la falta de integración social, influyen considerablemente en la retención de los estudiantes. Factores personales y demográficos, como la edad, el estado civil, la educación de los padres y el apoyo familiar, también desempeñan un papel importante en el abandono, especialmente entre aquellos con mayores responsabilidades familiares. A nivel institucional, la calidad de la gestión educativa y la

disponibilidad de instalaciones adecuadas son esenciales para una experiencia de aprendizaje favorable, al igual que la participación activa en actividades académicas, incluida la interacción en plataformas virtuales. En conjunto, estos factores reflejan la complejidad del fenómeno de deserción y la necesidad de estrategias integrales que atiendan estos múltiples aspectos.

Tabla 7 Factores que contribuyen al abandono escolar

Estudio	Factores
(Jiménez-Gutiérrez et al., 2024)	Dificultades para aprender, exclusión y vulnerabilidad de los estudiantes, problemas económicos y prejuicios
(Okoye et al., 2024)	El promedio de calificaciones de la escuela secundaria previa y la puntuación de ingreso/admisión de los estudiantes, demostró ser más eficaz para predecir la retención de los estudiantes en comparación con el uso de los datos de graduación.
(Delen et al., 2023)	Promedio de calificaciones, la cantidad de créditos aprobados, problemas financieros, ingresos bajos.
(Vaarma & Li, 2024)	Créditos acumulados, la cantidad de cursos reprobados y el conteo de actividades de Moodle.
(Krüger et al., 2023)	Dificultades financieras, calificaciones bajas y dificultades en diversas asignaturas, desmotivación y falta de interés en continuar con la educación.
(Porras et al., 2023)	Número de discusiones vistas en las plataformas, exámenes que no están completados y horas dedicadas al curso son indicadores importantes.
(Matz et al., 2023)	Estudiantes con dificultades financieras, un bajo rendimiento académico previo, medido por el GPA o las puntuaciones de exámenes, la falta de integración social y participación en la vida universitaria.
(Villegas-Ch et al., 2023)	Estudiantes con bajas calificaciones, baja asistencia, situación económica y antecedentes familiares.
(Mduma, 2023)	Fuente de ingresos del hogar y el tamaño del hogar, relación de alumnos con maestro, falta de instalaciones adecuadas.

(Tito et al., 2023)	Dificultades emocionales, baja autoestima, problemas de salud, dificultades financieras, estado civil, expectativas no cumplidas, falta de orientación profesional y el impacto de la COVID-19.
(Fauszt et al., 2023)	Antecedentes familiares, conexiones sociales, empleo, motivación, habilidades académicas, estrés, y dificultades financieras.
(Villarreal-Torres et al., 2023)	Rendimiento académico, materias reprobadas, repetición de año, desempeño docente, estrés de la persona, rendimiento en pregrado, el número de hijos, la motivación, la situación económica, el trabajo relacionado con su carrera, financiamiento de sus estudios.
(Segura et al., 2022)	Menor tiempo dedicado al estudio por semana, rendimiento académico del primer semestre, la preferencia por el curso, nivel educativo de los padres, la edad y el tipo de beca.
(Mnyawami et al., 2022)	Calificaciones, edad, distancia a la escuela, número de hijos, educación de los padres, género del estudiante, y medios para ir a la escuela.
(Niyogisubizo et al., 2022)	El bienestar social, el rendimiento del estudiante, las condiciones de aprendizaje, la edad, el género, el estado familiar, el patrocinio del estudiante, etc.
(Opazo et al., 2021)	Factores económicos individuales y familiares, factores institucionales (gestión, intervención institucional y seguimiento del estudiante), factores psicológicos (apoyo psicosocial y familiar), factores personales (motivación y relaciones sociales), así como factores académicos y demográficos.
(Kabathova & Drlik, 2021)	Número insuficiente de puntos obtenidos, tareas no entregadas o retrasadas, y baja actividad en el curso.
(Freitas et al., 2020)	Ingresos familiares bajos y dificultades financieras, edad, género, raza y procedencia del bachillerato, tamaño de la universidad y características socioeconómicas.
(Palacios et al., 2021)	Puntaje educativo secundario, índice de pobreza en la comunidad.

(Kuz & Morales, 2023) Promedio del estudiante en el primer período y el porcentaje de la beca.

La Tabla 8, presenta una visión general de diversos estudios sobre el uso de técnicas de Machine Learning para predecir la deserción escolar y mejorar la retención estudiantil. En general, se observa que estos modelos son efectivos para identificar a los estudiantes en riesgo de abandonar sus estudios, lo que permite a las instituciones educativas implementar intervenciones oportunas y personalizadas, como tutorías y apoyo financiero, aumentando así las posibilidades de permanencia. Los estudios destacan la importancia de utilizar conjuntos de datos amplios y longitudinales, así como la adaptabilidad de los modelos, ya sean individuales o centralizados, para abordar las necesidades específicas de los estudiantes. Asimismo, se enfatiza que la detección temprana del riesgo de abandono no solo ayuda a mejorar las tasas de retención, sino que también permite a las autoridades educativas tomar decisiones informadas, promoviendo un enfoque proactivo ante el problema de la deserción escolar.

Tabla 8 Impacto del machine learning en la reducción del abandono escolar

Estudio	Impacto
(Jiménez-Gutiérrez et al., 2024)	Cada una de las técnicas de Machine Learning implementadas, permite identificar la probabilidad de que un joven esté en riesgo de desertar, por causas ajenas a su voluntad, que se encuentre en situación de vulnerabilidad para continuar sus estudios en nivel medio superior o superior, y qué instituciones federales y educativas pueden brindar apoyo para su mejora.
(Okoye et al., 2024)	El modelo Bagging, con validación cruzada de 10 veces, es eficaz para predecir la retención (90.9%) y graduación (82.2%) de estudiantes. Su alto rendimiento permite identificar a quienes están en riesgo de abandono escolar de manera oportuna.
(Delen et al., 2023)	El modelo mejora el rendimiento predictivo al utilizar un conjunto de datos longitudinales más amplio y técnicas avanzadas de aprendizaje automático. Para luego ofrecer a responsables de políticas y administradores escolares un modelo de "caja negra" que permite tomar

(Vaarma & Li, 2024)	<p>decisiones informadas tanto a nivel colectivo como individual, facilitando así la retención de estudiantes.</p> <p>Ayuda a predecir el abandono escolar y por ello las universidades pueden intervenir antes de que los estudiantes abandonen, aumentando las posibilidades de retención. Finalmente proporcionan ideas prácticas sobre cuándo y cómo intervenir para evitar que los estudiantes abandonen, mejorando así las tasas de retención.</p>
(Krüger et al., 2023)	<p>El modelo permite detectar posibles abandonos escolares en diferentes momentos del año, lo que facilita la intervención temprana. Con la información obtenida, las instituciones pueden ofrecer apoyo personalizado, como tutorías o descuentos, para mantener a los estudiantes en la escuela.</p>
(Porras et al., 2023)	<p>Este estudio muestra que, para optimizar la predicción del abandono escolar, las instituciones educativas pueden elegir entre modelos individuales o centralizados según el tamaño y la naturaleza de sus datos. Los modelos individuales ofrecen un ajuste preciso para entidades pequeñas, aunque requieren más esfuerzo de mantenimiento, mientras que los modelos centralizados, al combinar datos de múltiples instituciones, permiten crear modelos robustos y generalizables, aunque presentan desafíos en privacidad y sesgos.</p>
(Matz et al., 2023)	<p>Permite la identificación temprana de estudiantes en riesgo y eso genera que haya intervenciones oportunas basadas en evidencia.</p>
(Villegas-Ch et al., 2023)	<p>Permite la identificación temprana, de esa manera las instituciones pueden diseñar estrategias de apoyo específicas para cada estudiante, aumentando la efectividad de las intervenciones.</p>
(Mduma, 2023)	<p>Los resultados permiten a las instituciones educativas anticipar y prevenir el abandono escolar mediante intervenciones tempranas y específicas.</p>

(Tito et al., 2023)	Ayuda a predecir qué estudiantes tienen mayor riesgo de abandonar, las universidades pueden intervenir a tiempo con apoyo académico y personal. Permite a las instituciones enfocar sus recursos en los estudiantes que más lo necesitan, mejorando la eficiencia de los programas de retención.
(Fauszt et al., 2023)	Ayuda a identificar a los estudiantes en riesgo lo antes posible para implementar intervenciones efectivas y reducir las tasas de abandono. El modelo es adecuado para la predicción de deserción estudiantil, proporcionando alta precisión, sensibilidad y especificidad, además de
(Villarreal-Torres et al., 2023)	ofrecer ventajas como la capacidad de manejar datos desbalanceados y mejorar los resultados mediante la generación automática de modelos de aprendizaje
(Segura et al., 2022)	El modelo identifica a los estudiantes en riesgo de abandono desde el primer semestre permitiendo a las instituciones intervenir de manera oportuna con estrategias de apoyo personalizadas, como tutorías y asesorías académicas.
(Mnyawami et al., 2022)	El modelo permite identificar con precisión a los estudiantes en riesgo de abandonar la escuela y de esa manera las instituciones pueden implementar estrategias proactivas y personalizadas para apoyar a estos estudiantes antes de que abandonen la escuela.
(Niyogisubizo et al., 2022)	La predicción del riesgo de abandono escolar motiva a los estudiantes a enfocarse en sus estudios y permite a las autoridades escolares tomar decisiones informadas, como promociones o exámenes adicionales, mientras que los maestros pueden intervenir tempranamente ante conductas de riesgo
(Opazo et al., 2021)	El modelo es muy útil para predecir la deserción estudiantil y pueden ayudar a las instituciones a identificar estudiantes en riesgo y mejorar las tasas de retención, beneficiando su viabilidad económica.

(Kabathova & Drlik, 2021)	El modelo permite predecir el abandono temprano del curso, facilitando intervenciones oportunas para mejorar la participación estudiantil y reducir la deserción mediante soluciones adaptadas a las necesidades de los estudiantes.
(Freitas et al., 2020)	El modelo permite identificar y clasificar a estudiantes en riesgo de abandono escolar según sus datos socioeconómicos del formulario de preinscripción, facilitando una evaluación preliminar antes de iniciar el curso. Estas características son importantes porque permiten una evaluación preliminar antes de que el estudiante comience el curso.
(Palacios et al., 2021)	El modelo facilita la predicción del nivel de deserción, lo que permite a las instituciones implementar medidas preventivas y mejorar la retención de estudiantes.
(Kuz & Morales, 2023)	La Ciencia de Datos Educativos y la aplicación de técnicas de aprendizaje automático permiten investigar las posibilidades de permanencia de los alumnos, lo que a su vez facilita la implementación de estrategias oportunas.

DISCUSIÓN DE RESULTADOS

El análisis del papel del machine learning como herramienta innovadora en la identificación, predicción y abordaje del abandono escolar resalta la convergencia entre la inteligencia artificial (IA) y la educación. La IA, como un campo amplio dentro de la ciencia de la computación, se dedica a crear máquinas capaces de realizar tareas que normalmente requieren inteligencia humana. Los sistemas de IA procesan grandes volúmenes de datos de entrenamiento etiquetados, buscando patrones que les permitan hacer pronósticos sobre estados futuros (Nimbalkar & Berad, 2021). Esta capacidad de análisis de datos es fundamental en la educación, donde el machine learning (ML) ha emergido como una herramienta relevante para mejorar diversos aspectos del aprendizaje y la retención estudiantil.

La inclusión del machine learning en el ámbito educativo ofrece un potencial significativo para abordar la deserción escolar, un fenómeno que representa un desafío global crítico. Según diversos estudios, las tasas

de abandono escolar son alarmantemente altas, afectando no solo a los individuos, sino también al desarrollo socioeconómico de las comunidades (Hassan et al., 2024; Selim & Rezk, 2023; Vaarma & Li, 2024). En este contexto, el machine learning permite modelar sistemas que no solo identifican a los estudiantes en riesgo de abandono, sino que también analizan factores que contribuyen a este comportamiento, lo que facilita la creación de intervenciones personalizadas.

La identificación de los modelos de machine learning más efectivos en la predicción del abandono escolar es un paso crucial para desarrollar intervenciones adecuadas y efectivas en el ámbito educativo. Según los estudios revisados, se ha evidenciado que varios modelos destacan en su capacidad para predecir la deserción escolar, cada uno con sus propias características y ventajas.

En primer lugar, las Redes Neuronales Artificiales (ANN) emergen como uno de los modelos más eficaces (Jiménez-Gutiérrez et al., 2024; Delen et al., 2023; Vaarma & Li, 2024, entre otros). Este modelo es conocido por su capacidad para manejar grandes volúmenes de datos y por su eficacia en la identificación de patrones complejos. Su versatilidad y capacidad de aprendizaje a partir de datos no lineales la convierten en una opción preferida en muchos estudios sobre abandono escolar.

Los Árboles de Decisión (DT) también se destacan, siendo mencionados en diversos estudios y reconocidos por su facilidad de interpretación y su capacidad para modelar relaciones no lineales entre variables (Vaarma & Li, 2024; Matz et al., 2023; Fauszt et al., 2023, entre otros). Este enfoque es especialmente útil en entornos educativos donde las decisiones deben ser claras y comprensibles para los educadores y administradores.

Por otro lado, la Máquina de Soporte de Vectores (SVM) y los Vecinos Más Cercanos (KNN) también aparecen con frecuencia. SVM es especialmente útil en la clasificación de datos en entornos de alta dimensionalidad, lo que lo hace ideal para abordar el abandono escolar al considerar múltiples factores que pueden influir en la deserción. KNN, por su parte, es conocido por su simplicidad y efectividad en la identificación de similitudes entre estudiantes, lo que puede ser crucial para detectar patrones de abandono. Adicionalmente, la Regresión Logística (LR) y Random Forest (RF) son otros algoritmos prominentes. La regresión logística es ampliamente utilizada en la predicción de eventos binarios, como la deserción escolar, gracias a su interpretación directa de probabilidades. Random Forest, por su parte, ofrece robustez y resistencia al sobreajuste, siendo eficaz en la clasificación y regresión.

Los resultados obtenidos analizadas también reflejan la diversidad de algoritmos utilizados en la predicción del abandono escolar. La Red Neuronal Artificial (ANN) y los Árboles de Decisión (DT) se encuentran entre los algoritmos más populares, mientras que la Máquina de Soporte de Vectores (SVM) y la Regresión Logística (LR) son igualmente comunes. Esta amplia variedad sugiere que no existe un enfoque único para abordar el abandono escolar, sino que la combinación de múltiples métodos puede ofrecer la mejor perspectiva y resultados.

Además, es relevante destacar que la utilización de modelos híbridos, como se menciona en los estudios de Niyogisubizo et al. (2022), representa una tendencia creciente en la investigación. Estos modelos combinan diferentes algoritmos para mejorar la precisión y efectividad de las predicciones, lo que puede ser un enfoque prometedor para abordar la complejidad del abandono escolar.

Por consiguiente, la identificación de los modelos de machine learning más efectivos en la predicción del abandono escolar es un área de investigación en constante evolución, con la capacidad de transformar la forma en que las instituciones educativas abordan este fenómeno. La combinación de diferentes técnicas y la implementación de modelos avanzados pueden facilitar la identificación de estudiantes en riesgo y la creación de intervenciones personalizadas, contribuyendo a la retención estudiantil y al éxito académico. Esta discusión subraya la importancia de seguir investigando y aplicando estos modelos en contextos educativos para optimizar las estrategias de intervención y mejorar la calidad de la educación.

La investigación sobre el abandono escolar es un tema crítico que abarca múltiples dimensiones y factores que influyen en la decisión de los estudiantes de continuar o no su educación. Al examinar los factores que contribuyen al abandono escolar, se puede observar una interrelación compleja entre variables académicas, económicas, sociales y emocionales, que, al integrarse en modelos de machine learning, pueden proporcionar información valiosa para predecir y mitigar este fenómeno.

Los estudios revisados destacan que las dificultades académicas y financieras son recurrentemente citadas como factores determinantes. Por ejemplo, Jiménez-Gutiérrez et al. (2024) identifican la vulnerabilidad de los estudiantes y problemas económicos como contribuyentes significativos al abandono escolar. De manera similar, Krüger et al. (2023) señalan que las calificaciones bajas, junto con dificultades en varias asignaturas, afectan la motivación y el interés de los estudiantes en continuar sus estudios. Estos factores

académicos son críticos, ya que los modelos de machine learning pueden ser entrenados para reconocer patrones en el rendimiento académico que correlacionan con la deserción escolar.

Además, el promedio de calificaciones previas y las puntuaciones de ingreso son esenciales para la predicción del abandono. Según Okoye et al. (2024), estos indicadores demuestran ser más eficaces para predecir la retención que los datos de graduación. Esto sugiere que la inclusión de variables académicas tempranas en los modelos puede mejorar significativamente su capacidad predictiva.

Otros factores como el bienestar social, las condiciones de aprendizaje, y la relación con los maestros también se mencionan frecuentemente en la literatura. Mduma (2023) enfatiza la influencia del tamaño del hogar y la fuente de ingresos del hogar, mientras que Mnyawami et al. (2022) destacan cómo la distancia a la escuela y el nivel educativo de los padres impactan la decisión de un estudiante de permanecer en la escuela. Estos factores sociales y demográficos, si se integran adecuadamente en los modelos de machine learning, pueden ofrecer una visión más holística de los estudiantes en riesgo de abandono.

Los estudios también mencionan factores emocionales y psicológicos como la autoestima, la desmotivación y las dificultades de salud, que afectan la permanencia de los estudiantes en el sistema educativo (Tito et al., 2023). Estos factores pueden ser difíciles de cuantificar, pero los enfoques de machine learning, particularmente aquellos que utilizan redes neuronales, pueden ser efectivos en la identificación de patrones en datos cualitativos que podrían correlacionarse con la deserción escolar.

En cuanto a la integración de estos factores en modelos de machine learning, la diversidad de variables sugiere que un enfoque de múltiples factores es esencial. Por ejemplo, los modelos híbridos que combinan algoritmos como Random Forest y Redes Neuronales pueden permitir una mejor comprensión de las interacciones entre los diferentes factores. Niyogisubizo et al. (2022) apuntan a la importancia de combinar factores económicos, académicos, y psicológicos en los modelos predictivos, lo que puede facilitar el desarrollo de intervenciones más efectivas y personalizadas para los estudiantes en riesgo.

Por lo tanto, la examinación de los factores que contribuyen al abandono escolar revela la complejidad del fenómeno y la necesidad de un enfoque multidimensional en su estudio. Integrar estos factores en modelos de machine learning no solo permite predecir el abandono escolar con mayor precisión, sino que también proporciona a los educadores y administradores herramientas para identificar a los estudiantes en riesgo y desarrollar estrategias de intervención más efectivas.

La implementación de técnicas de machine learning (ML) en el ámbito educativo ofrece un enfoque innovador para abordar el problema del abandono escolar, al permitir a las instituciones educativas desarrollar estrategias de intervención más eficaces y específicas. Diversos estudios destacan cómo estas técnicas no solo identifican a los estudiantes en riesgo, sino que también facilitan la formulación de intervenciones adaptadas a sus necesidades.

Por ejemplo, Jiménez-Gutiérrez et al. (2024) enfatizan que las técnicas de ML pueden identificar con precisión la probabilidad de que un estudiante se encuentre en situación de riesgo, lo que permite a las instituciones no solo intervenir a tiempo, sino también conectar a los estudiantes con recursos de apoyo. Asimismo, el modelo de Okoye et al. (2024), que emplea la técnica de Bagging, muestra una notable eficacia en la predicción de la retención y graduación, lo que permite a las universidades intervenir antes de que se produzca la deserción. Esto resalta la importancia de anticipar y actuar proactivamente, en lugar de reaccionar ante situaciones ya críticas.

Además, Delen et al. (2023) aportan la idea de que el uso de un conjunto de datos longitudinales amplios y técnicas avanzadas mejora el rendimiento predictivo, proporcionando a los administradores educativos un modelo de "caja negra". Este modelo permite decisiones informadas tanto a nivel colectivo como individual, lo que es esencial para abordar el abandono escolar de manera efectiva. La capacidad de los modelos para ofrecer información detallada sobre el comportamiento de los estudiantes permite que las instituciones puedan diseñar estrategias de apoyo personalizadas, como tutorías o asesorías académicas, tal como se menciona en el estudio de Segura et al. (2022).

Por otro lado, estudios como el de Porras et al. (2023) sugieren que las instituciones deben considerar el tamaño y la naturaleza de sus datos al elegir entre modelos individuales o centralizados, cada uno con sus ventajas y desventajas en términos de precisión y mantenimiento. Esto es crucial, ya que las decisiones sobre qué modelo implementar pueden influir en la efectividad de las intervenciones diseñadas para prevenir el abandono.

La capacidad de ML para detectar abandonos en diferentes momentos del año, como se indica en Krüger et al. (2023), permite intervenciones más oportunas. La identificación temprana no solo beneficia a los estudiantes, sino que también optimiza el uso de recursos por parte de las instituciones, permitiendo que se concentren en aquellos que más lo necesitan, como se describe en Tito et al. (2023).

Finalmente, la investigación de Matz et al. (2023) concluye que la identificación anticipada de estudiantes en riesgo fomenta intervenciones basadas en evidencia, lo que contribuye a aumentar la efectividad de las estrategias implementadas. Este enfoque proactivo, basado en datos y análisis predictivo, transforma la forma en que las instituciones educativas enfrentan el desafío del abandono escolar, haciéndolas más resilientes y capaces de adaptarse a las necesidades cambiantes de sus estudiantes.

En definitiva, la implementación de machine learning en la educación no solo contribuye a la identificación temprana de estudiantes en riesgo de abandono escolar, sino que también permite a las instituciones educativas desarrollar estrategias de intervención más informadas y personalizadas, aumentando así las posibilidades de retención y éxito académico de los estudiantes.

CONCLUSIONES

El análisis del machine learning (ML) como herramienta innovadora en la educación destaca su potencial transformador en la identificación y prevención del abandono escolar. Los estudios revisados demuestran que diversas técnicas de ML permiten a las instituciones educativas predecir con alta precisión los riesgos de deserción e implementar intervenciones personalizadas que abordan las necesidades específicas de los estudiantes en riesgo.

Entre los modelos más efectivos, las Redes Neuronales Artificiales (ANN) han mostrado un sólido respaldo en la literatura reciente, lo que sugiere su capacidad para manejar la complejidad de los datos y realizar predicciones precisas sobre la deserción estudiantil. Otros modelos, como los Vecinos Más Cercanos (KNN), la Regresión Lineal (LR) y los Árboles de Decisión (DT), también han demostrado eficacia en la identificación de estudiantes en riesgo, facilitando intervenciones oportunas.

Los estudios indican que el rendimiento académico previo, medido a través de calificaciones y créditos acumulados, es un predictor clave del abandono escolar, lo que resalta la importancia de la preparación educativa de los estudiantes. Además, factores como las dificultades financieras, la situación familiar y la falta de apoyo social son determinantes, evidenciando cómo el contexto personal y socioeconómico influye en su capacidad para permanecer en el sistema educativo.

La implementación de machine learning, a través de modelos predictivos como el Bagging, permite a las instituciones identificar con precisión a los estudiantes en riesgo de deserción, propiciando intervenciones tempranas y efectivas.

Se recomienda que las instituciones no solo adopten estas tecnologías, sino que también inviertan en la capacitación del personal en el uso y análisis de herramientas de ML. La colaboración entre educadores, administradores y expertos en ciencia de datos será esencial para optimizar las estrategias de intervención y reducir el abandono escolar. En última instancia, el machine learning se establece como una herramienta clave para mejorar la calidad educativa y garantizar que todos los estudiantes tengan la oportunidad de completar su formación académica.

REFERENCIAS BIBLIOGRÁFICAS

- Alenezi, H. S., & Faisal, M. H. (2020). Utilizing crowdsourcing and machine learning in education: Literature review. *Education and Information Technologies*, 25(4), 2971–2986.
<https://doi.org/10.1007/s10639-020-10102-w>
- Alhabeeb, S., Alrusayni, N., Almutiri, R., Alhumud, S., & Al-Hagery, M. A. (2024). Blockchain and machine learning in education: a literature review. *IAES International Journal of Artificial Intelligence*, 13(1), 581–596. <https://doi.org/10.11591/ijai.v13.i1.pp581-596>
- Bettany-Saltikov, J., & McSherry, R. (2024). *How to do a Systematic Literature Review in Nursing: A Step-by-Step Guide* (3era. Ed. (ed.)). <https://goo.su/wMxVzm>
- Delen, D., Davazdahemami, B., & Rasouli Dezfouli, E. (2023). Predicting and Mitigating Freshmen Student Attrition: A Local-Explainable Machine Learning Framework. *Information Systems Frontiers*, 26(2), 641–662. <https://doi.org/10.1007/s10796-023-10397-3>
- European Commission. (2024). *Reducir el abandono escolar prematuro: tratamiento y prevención del abandono escolar*. <https://school-education.ec.europa.eu/mk/learn/courses/reducing-early-school-leaving-treatment-and-prevention-dropouts>
- Fauszt, T., Erdélyi, K., Dobák, D., Bognár, L., & Kovács, E. (2023). Design of a Machine Learning Model to Predict Student Attrition. *International Journal of Emerging Technologies in Learning*, 18(17), 184–195. <https://doi.org/10.3991/ijet.v18i17.41449>
- Freitas, F. A. d. S., Vasconcelos, F. F. X., Peixoto, S. A., Hassan, M. M., Ali Akber Dewan, M., de Albuquerque, V. H. C., & Rebouças Filho, P. P. (2020). IoT system for school dropout prediction using machine learning techniques based on socioeconomic data. *Electronics*, 9, 1–14.
<https://doi.org/10.3390/electronics9101613>

- Gómez-Pulido, J. A., Park, Y., Soto, R., & Lanza-Gutiérrez, J. M. (2023). Data Analytics and Machine Learning in Education. *Applied Sciences (Switzerland)*, *13*(3), 13–15.
<https://doi.org/10.3390/app13031418>
- Hassan, M. A., Muse, A. H., & Nadarajah, S. (2024). applied sciences Learning : Insights from the 2022 National Education Accessibility Survey in Somaliland. *Applied Sciences*, *14*.
<https://doi.org/10.3390/app14177593>
- Jiménez-Gutiérrez, A. L., Mota-Hernández, C. I., Mezura-Montes, E., & Alvarado-Corona, R. (2024). Application of the performance of machine learning techniques as support in the prediction of school dropout. *Scientific Reports*, *14*, 1–8. <https://doi.org/10.1038/s41598-024-53576-1>
- Kabathova, J., & Drlik, M. (2021). Towards predicting student’s dropout in university courses using different machine learning techniques. *Applied Sciences*, *11*, 1–19.
<https://doi.org/10.3390/app11073130>
- Krüger, J. G. C., Britto, A. de S., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications*, *233*, 1–9.
<https://doi.org/10.1016/j.eswa.2023.120933>
- Kuz, A., & Morales, R. (2023). Ciencia de Datos Educativos y aprendizaje automático: un caso de estudio sobre la deserción estudiantil universitaria en México. *Education in the Knowledge Society*, *24*, 1–14. <https://doi.org/10.14201/eks.30080>
- Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., Dinu, A., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Scientific Reports*, *13*(1), 1–16. <https://doi.org/10.1038/s41598-023-32484-w>
- Mduma, N. (2023a). Data Balancing Techniques for Predicting Student Dropout Using Machine Learning. *Data*, *8*(3). <https://doi.org/10.3390/data8030049>
- Mduma, N. (2023b). Data Balancing Techniques for Predicting Student Dropout Using Machine Learning. *Data*, *8*(49), 1–14. <https://doi.org/10.3390/data8030049>
- Mnyawami, Y. N., Maziku, H. H., & Mushi, J. C. (2022). Enhanced Model for Predicting Student Dropouts in Developing Countries Using Automated Machine Learning Approach: A Case of Tanzanian’s Secondary Schools. *Applied Artificial Intelligence*, *36*(1).

<https://doi.org/10.1080/08839514.2022.2071406>

- Nimbalkar, A. A., & Berad, A. T. (2021). The increasing importance of AI applications in E-Commerce. *Vidyabharati International Interdisciplinary Research Journal*, 13(1), 67–77. <https://www.viirj.org/vol13issue1/56.pdf>
- Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, 3, 1–12. <https://doi.org/10.1016/j.caeai.2022.100066>
- Okagbue, E. F., Ezeachikulo, U. P., Akintunde, T. Y., Tsakuwa, M. B., Ilokanulo, S. N., Obiasoanya, K. M., Ilodibe, C. E., & Ouattara, C. A. T. (2023). A comprehensive overview of artificial intelligence and machine learning in education pedagogy: 21 Years (2000–2021) of research indexed in the scopus database. *Social Sciences and Humanities Open*, 8(1), 100655. <https://doi.org/10.1016/j.ssaho.2023.100655>
- Okoye, K., Njanji, J. T., Escamilla, J., & Hosseini, S. (2024a). Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education. *Computers and Education: Artificial Intelligence*, 6, 100205. <https://doi.org/10.1016/j.caeai.2024.100205>
- Okoye, K., Njanji, J. T., Escamilla, J., & Hosseini, S. (2024b). Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education. *Computers and Education: Artificial Intelligence*, 6, 1–13. <https://doi.org/10.1016/j.caeai.2024.100205>
- Opazo, D., Moreno, S., Álvarez-Miranda, E., & Pereira, J. (2021). Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics*, 9, 1–27. <https://doi.org/10.3390/math9202599>
- Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*, 23, 1–23. <https://doi.org/10.3390/e23040485>
- Porras, M., Lara, J. A., Romero, C., & Ventura, S. (2023). A Case-Study Comparison of Machine Learning Approaches for Predicting Student's Dropout from Multiple Online Educational Entities. *Algorithms*, 16(554), 1–21. <https://doi.org/10.3390/a16120554>



- Razaulla, S. M., Pasha, M., & Farooq, M. U. (2022). *Integration of Machine Learning in Education: Challenges, Issues and Trends BT - Machine Learning and Internet of Things for Societal Issues* (C. Satyanarayana, X.-Z. Gao, C.-Y. Ting, & N. B. Muppalaneni (eds.); pp. 23–34). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-5090-1_2
- Sahana, S., Singh, D., & Nath, I. (2023). Importance of AI and ML Towards Smart Sensor Network Utility Enhancement. *Encyclopedia of Data Science and Machine Learning*, 240–262. <https://doi.org/10.4018/978-1-7998-9220-5.ch015>
- Segura, M., Mello, J., & Hernández, A. (2022). Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics*, 10(18), 1–20. <https://doi.org/10.3390/math10183359>
- Selim, K. S., & Rezk, S. S. (2023). On predicting school dropouts in Egypt: A machine learning approach. *Education and Information Technologies*, 28(7), 9235–9266. <https://doi.org/10.1007/s10639-022-11571-x>
- Shah, D., Patel, D., Adesara, J., Hingu, P., & Shah, M. (2021). Exploiting the Capabilities of Blockchain and Machine Learning in Education. *Augmented Human Research*, 6(1), 1. <https://doi.org/10.1007/s41133-020-00039-7>
- Sinha, A., Menon, G. R., & John, D. (2022). *Beginner's guide for systematic reviews*. https://main.icmr.nic.in/sites/default/files/upload_documents/BEGINNERS_GUIDE_FINAL_BO OK.pdf
- Tito, A. E. A., Condori, B. O. H., & Vera, Y. P. (2023). Comparative analysis of Machine Learning Techniques for the prediction of cases of university dropout. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, 51(09), 84–98. <https://doi.org/10.17013/risti.51.84-98>
- Tiwari, R. (2023). The integration of AI and machine learning in education and its potential to personalize and improve student learning experiences. *Interantional Journal of Scientific Research in Engineering and Management*, 7(2). <https://doi.org/10.55041/ijrsrem17645>
- Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society*, 76, 1–10. <https://doi.org/10.1016/j.techsoc.2024.102474>

- Villar, A., & Velini, C. R. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. *Discover Artificial Intelligence*, 4(2). <https://doi.org/10.1007/s44163-023-00079-z>
- Villarreal-Torres, H., Ángeles-Morales, J., Cano-Mejía, J., Mejía-Murillo, C., Flores-Reyes, G., Palomino-Márquez, M., Marín-Rodríguez, W., & Andrade-Girón, D. (2023). Classification model for student dropouts using machine learning: A case study. *EAI Endorsed Transactions on Scalable Information Systems*, 10(5), 1–12. <https://doi.org/10.4108/eetsis.vi.3455>
- Villegas-Ch, W., Govea, J., & Revelo-Tapia, S. (2023). Improving Student Retention in Institutions of Higher Education through Machine Learning: A Sustainable Approach. *Sustainability (Switzerland)*, 15(19), 1–20. <https://doi.org/10.3390/su151914512>
- Wu, J. (2020). Machine Learning in Education. *2020 International Conference on Modern Education and Information Management (ICMEIM)*, 56–63. <https://doi.org/10.1109/ICMEIM51375.2020.00020>
- Zhou, Y., & Song, Z. (2020). Effectiveness analysis of machine learning in education big data. *Journal of Physics: Conference Series*, 1651(1). <https://doi.org/10.1088/1742-6596/1651/1/012105>