





Modelo predictivo del riesgo de abandono escolar en educación básica: análisis de factores socioeconómicos y rendimiento académico en niños de 7 a 11 años

Predictive Model of School Dropout Risk in Basic Education: Analysis of Socioeconomic Factors and Academic Performance in Children Aged 7 to 11

Angelica Tatiana Rosado Quinto ¹[\[0000-0002-2707-9666\]](#)  Mercy Yalinda Tigua Quimiz ²[\[0009-0003-0424-1733\]](#) 
Katty Pamela Bermeo Erazo ³[\[0009-0002-7425-6730\]](#)  Mirian Yessenia Torres Salinas ⁴[\[0009-0009-9833-3631\]](#) 

Ministerio de Educacion Deporte y Cultura, MINEDEC.

mercy.tigua@educacion.gob.ec katty.bermeo@docentes.educacion.edu.ec
yessenia.torres@educacion.gob.ec

CITA EN APA:

Rosado Quinto, A. T., Tigua Quimiz, M. Y., Bermeo Erazo, K. P., & Torres Salinas, M. Y. (2026). Modelo predictivo del riesgo de abandono escolar en educación básica: análisis de factores socioeconómicos y rendimiento académico en niños de 7 a 11 años. *Tesla Revista Científica*, 6(1).

Recibido: 2026-02-17

Aceptado: 2026-04-04

Publicado: 2026-05-11

TESLA

Revista Científica

ISSN: 2796-9320



Los contenidos de este artículo están bajo una licencia de Creative Commons Attribution 4.0 International (CC BY 4.0)

Los autores conservan los derechos morales y patrimoniales de sus obras. The contents of this article are under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. The authors retain the moral and patrimonial rights of their works.

Resumen. La deserción escolar constituye un fenómeno multidimensional asociado a factores académicos, socioeconómicos, familiares, institucionales y contextuales. El presente artículo tiene como objetivo analizar la literatura científica reciente sobre la aplicación de modelos predictivos, regresión logística y técnicas de *machine learning* en la predicción del abandono escolar. La investigación se desarrolló mediante una revisión bibliográfica narrativa con orientación sistemática, considerando estudios publicados en bases académicas relacionadas con educación, analítica del aprendizaje e inteligencia artificial. Los hallazgos evidencian que la regresión logística destaca por su interpretabilidad, mientras que algoritmos como Random Forest, XGBoost, redes neuronales y máquinas de soporte vectorial presentan alto potencial predictivo. Se concluye que estos modelos pueden fortalecer la detección temprana del riesgo de deserción, siempre que sean aplicados con criterios éticos, explicabilidad y orientación pedagógica.

Palabras Clave: deserción escolar; modelos predictivos; regresión logística; *machine learning*; inteligencia artificial; analítica educativa.

Abstract: School dropout is a multidimensional phenomenon associated with academic, socioeconomic, family, institutional, and contextual factors. This article aims to analyze recent scientific literature on the application of predictive models, logistic regression, and machine learning techniques for predicting school dropout. The study was conducted through a narrative literature review with a systematic orientation, considering studies published in academic databases related to education, learning analytics, and artificial intelligence. The findings show that logistic regression stands out for its interpretability, while algorithms such as Random Forest, XGBoost, neural networks, and support vector machines show high predictive potential. It is concluded that these models can strengthen early detection of dropout risk, provided they are applied with ethical criteria, explainability, and pedagogical orientation.

Keywords: school dropout; predictive models; logistic regression; machine learning; artificial intelligence; educational analytics.

1 INTRODUCCIÓN

La deserción escolar constituye uno de los principales desafíos de los sistemas educativos contemporáneos debido a sus repercusiones sociales, económicas y académicas. Este fenómeno afecta directamente el desarrollo integral de los estudiantes, limita sus oportunidades futuras y profundiza las desigualdades sociales, especialmente en contextos vulnerables y rurales. De acuerdo con la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO] (2024), el abandono escolar continúa representando una barrera significativa para garantizar una educación inclusiva, equitativa y de calidad, particularmente en países en vías de desarrollo donde persisten brechas estructurales relacionadas con pobreza, acceso tecnológico y desigualdad educativa.

La deserción escolar es un fenómeno multicausal influenciado por factores académicos, familiares, socioeconómicos, emocionales e institucionales. Diversas investigaciones han evidenciado que variables como el bajo rendimiento académico, la repitencia, la falta de apoyo familiar, la violencia, las limitaciones económicas y el escaso acceso a recursos tecnológicos incrementan significativamente el riesgo de abandono escolar (Quispe, 2010; Vélez & Ruperti, 2024). Asimismo, estudios recientes señalan que los factores socioemocionales, la motivación académica y el contexto escolar desempeñan un papel fundamental en la permanencia estudiantil y el éxito educativo (Bustamante Neira & Cabrera Berrezueta, 2022).

En América Latina, esta problemática adquiere una relevancia especial debido a las profundas desigualdades sociales y educativas existentes entre sectores urbanos y rurales. La Comisión Económica para América Latina y el Caribe [CEPAL] (2023) sostiene que las brechas socioeconómicas y digitales han incrementado las dificultades de acceso y permanencia escolar, especialmente después de la pandemia de COVID-19. En este contexto, países como Colombia, Ecuador y Perú continúan enfrentando importantes desafíos relacionados con la retención estudiantil y la implementación de estrategias efectivas para prevenir el abandono escolar.

Frente a esta realidad, en los últimos años se ha incrementado el interés científico por el uso de modelos predictivos y técnicas de inteligencia artificial aplicadas al ámbito educativo. Estas herramientas permiten identificar patrones asociados al riesgo de deserción escolar mediante el análisis de variables académicas, sociales y económicas, facilitando la detección temprana de estudiantes en situación de vulnerabilidad. Según Delgado Aguilar et al. (2014), la regresión logística constituye uno de los modelos estadísticos más utilizados en la predicción del rendimiento académico y abandono escolar debido a su capacidad para estimar probabilidades a partir de variables predictoras.

De igual manera, el desarrollo del machine learning y la minería de datos educativos ha generado nuevas perspectivas para el análisis de fenómenos académicos complejos. Modelos como árboles de decisión, Random Forest, redes neuronales artificiales y Support Vector Machine (SVM) han demostrado altos niveles de precisión en la clasificación de estudiantes con riesgo de abandono escolar (Bastidas Guacho et al., 2021). Estas herramientas permiten procesar grandes volúmenes de información y detectar relaciones complejas entre variables que, mediante métodos tradicionales, resultarían difíciles de identificar.

El avance de la inteligencia artificial aplicada a la educación ha transformado progresivamente la manera en que las instituciones educativas abordan los problemas relacionados con el rendimiento académico y la permanencia estudiantil. Actualmente, la literatura científica evidencia un creciente interés por desarrollar sistemas predictivos capaces de anticipar escenarios de riesgo y contribuir a la toma de decisiones basadas en evidencia (Álvarez-Yaulema et al., 2024). En este sentido, los modelos predictivos representan una alternativa relevante para fortalecer las estrategias de intervención temprana y reducir los índices de deserción escolar.

No obstante, a pesar del incremento de investigaciones relacionadas con modelos predictivos educativos, aún existe dispersión teórica y metodológica respecto a las variables más utilizadas, los algoritmos con mejores niveles de desempeño y las limitaciones presentes en los estudios desarrollados. Asimismo, muchas investigaciones se enfocan en contextos específicos o utilizan metodologías heterogéneas, dificultando la construcción de una visión integral sobre el uso de inteligencia artificial y machine learning en la predicción de la deserción escolar.

Resulta pertinente desarrollar una revisión bibliográfica que permita sistematizar y analizar críticamente la producción científica relacionada con la aplicación de modelos predictivos, regresión logística y técnicas de aprendizaje automático en el ámbito educativo. Según Hernández-Sampieri y Mendoza (2021), las revisiones bibliográficas constituyen una herramienta fundamental para identificar tendencias investigativas, analizar enfoques metodológicos y sintetizar el conocimiento científico existente sobre un fenómeno determinado.

Por ello, el presente artículo tiene como objetivo analizar la literatura científica relacionada con la aplicación de modelos predictivos y técnicas de machine learning en la predicción de la deserción escolar, mediante una revisión bibliográfica de investigaciones publicadas en bases de datos académicas indexadas. A través de este análisis, se busca identificar los principales modelos utilizados, las variables predictoras más relevantes y las tendencias actuales en el uso de inteligencia artificial aplicada a la educación.

La relevancia de esta investigación radica en que sus resultados pueden contribuir al fortalecimiento de futuras investigaciones y al diseño de estrategias educativas orientadas a la prevención de la deserción escolar mediante herramientas tecnológicas y modelos predictivos sustentados en evidencia científica. Además, este estudio aporta una visión actualizada sobre el estado del conocimiento en torno al uso de inteligencia artificial y machine learning en contextos educativos, favoreciendo el desarrollo de propuestas innovadoras para mejorar la permanencia estudiantil y la calidad educativa.

1.1 Justificación de la investigación

La presente investigación se justifica por la creciente importancia que han adquirido los modelos predictivos y las técnicas de inteligencia artificial dentro del ámbito educativo, especialmente en el análisis y prevención de la deserción escolar. En los últimos años, el abandono estudiantil se ha convertido en una problemática de interés internacional debido a sus implicaciones sociales, económicas y académicas, afectando tanto el desarrollo individual de los estudiantes como el progreso de los sistemas educativos y de las sociedades en general (UNESCO, 2024).

Desde una perspectiva educativa, la deserción escolar representa uno de los principales obstáculos para garantizar una educación inclusiva, equitativa y de calidad. Diversos estudios han demostrado que factores como el bajo rendimiento académico, la pobreza, la repitencia, la falta de apoyo familiar y las limitaciones tecnológicas influyen significativamente en la permanencia estudiantil (Quispe, 2010; Vélez & Ruperti, 2024). En este contexto, la necesidad de desarrollar estrategias de identificación temprana del riesgo de abandono escolar ha impulsado el interés científico por el uso de herramientas estadísticas, minería de datos y algoritmos de machine learning aplicados a la educación.

Desde el punto de vista teórico, esta revisión bibliográfica resulta relevante porque permite sistematizar y analizar críticamente la producción científica relacionada con la aplicación de modelos predictivos en la detección de riesgo de deserción escolar. La literatura existente evidencia un crecimiento significativo en investigaciones orientadas al uso de regresión logística, árboles de decisión, redes neuronales y otros algoritmos de aprendizaje automático para el análisis de fenómenos educativos complejos (Bastidas Guacho et al., 2021). Sin embargo, aún existe dispersión metodológica y conceptual respecto a las variables más utilizadas, los modelos con mayor precisión predictiva y las limitaciones presentes en los estudios desarrollados.

Asimismo, la investigación adquiere relevancia científica debido a que contribuye a consolidar el conocimiento existente sobre inteligencia artificial aplicada a la educación, un campo

que continúa expandiéndose como consecuencia de los avances tecnológicos y la transformación digital de los sistemas educativos. Según Álvarez-Yaulema et al. (2024), el análisis predictivo basado en datos representa una herramienta estratégica para fortalecer la toma de decisiones educativas y mejorar los mecanismos de intervención temprana en estudiantes en situación de vulnerabilidad académica.

Desde una perspectiva metodológica, el estudio resulta pertinente porque permite identificar tendencias investigativas, enfoques metodológicos predominantes y modelos predictivos más utilizados en la literatura científica reciente. Además, la revisión bibliográfica facilita la comparación de resultados obtenidos en distintos contextos educativos, contribuyendo a una comprensión más amplia del fenómeno de la deserción escolar y de las herramientas tecnológicas empleadas para su análisis. Hernández-Sampieri y Mendoza (2021) señalan que las revisiones bibliográficas permiten sintetizar información científica relevante y establecer nuevas líneas de investigación a partir del análisis crítico de estudios previos.

En el ámbito social, la investigación posee gran importancia debido a que la deserción escolar afecta principalmente a estudiantes pertenecientes a sectores vulnerables, profundizando las desigualdades educativas y limitando las oportunidades de desarrollo social y económico. La implementación de modelos predictivos y sistemas de inteligencia artificial puede contribuir significativamente a la prevención temprana del abandono escolar mediante la identificación oportuna de factores de riesgo y la generación de estrategias de intervención más eficientes (CEPAL, 2023).

Por otro lado, la revisión de literatura sobre machine learning y regresión logística aplicada a la educación resulta relevante para instituciones educativas, investigadores y responsables de políticas públicas, ya que proporciona evidencia científica actualizada sobre las principales herramientas tecnológicas utilizadas en la predicción del abandono escolar. Esta información puede servir como base para el diseño de programas educativos, estrategias institucionales y políticas orientadas a fortalecer la permanencia estudiantil y reducir los índices de deserción.

La presente investigación se justifica por la necesidad de contar con estudios actualizados que integren los avances recientes en inteligencia artificial, minería de datos y aprendizaje automático aplicados al contexto educativo. La rápida evolución de estas tecnologías exige una revisión constante de la literatura científica para comprender sus alcances, limitaciones y potencialidades dentro de los procesos educativos contemporáneos. En este sentido, el estudio aporta una visión integral sobre el estado actual de los modelos predictivos utilizados en la detección del riesgo de deserción escolar y contribuye al fortalecimiento del conocimiento científico en el campo de la analítica educativa.

1.2 Planteamiento del problema

En la actualidad, la deserción escolar continúa siendo una problemática educativa de alta relevancia social, debido a que compromete la continuidad de las trayectorias académicas, limita el desarrollo de capacidades humanas y reproduce condiciones de desigualdad estructural. Aunque los sistemas educativos han avanzado en cobertura, acceso y permanencia, el abandono escolar persiste como un fenómeno complejo que afecta con mayor intensidad a estudiantes en situación de vulnerabilidad económica, social, territorial y familiar. De acuerdo con la UNESCO (2024), garantizar una educación inclusiva y equitativa requiere no solo ampliar el acceso, sino también asegurar que los estudiantes permanezcan y culminen sus procesos formativos.

La literatura científica ha señalado que la deserción escolar no responde a una causa única, sino a la interacción de múltiples factores. Entre los elementos más recurrentes se encuentran el bajo rendimiento académico, la repitencia, la pobreza, la falta de acompañamiento familiar, las dificultades de acceso a recursos educativos, la violencia, el trabajo infantil, la desmotivación y las limitaciones institucionales para ofrecer respuestas oportunas a los estudiantes en riesgo (Rumberger, 2011; Tinto, 1993). Esta característica multidimensional dificulta la identificación temprana del abandono escolar y exige enfoques analíticos capaces de integrar variables académicas, socioeconómicas, familiares y contextuales.

En el contexto latinoamericano, el problema adquiere una dimensión particular debido a las brechas históricas de desigualdad que afectan a los sistemas educativos. La Comisión Económica para América Latina y el Caribe [CEPAL] (2023) ha señalado que las condiciones de pobreza, exclusión social, desigualdad territorial y brecha digital inciden directamente en las oportunidades educativas de niños, niñas y adolescentes. Estas condiciones se agudizaron durante y después de la pandemia de COVID-19, periodo en el que muchos estudiantes enfrentaron interrupciones en sus aprendizajes, dificultades de conectividad, reducción de ingresos familiares y pérdida de contacto con las instituciones educativas (UNESCO, UNICEF, & World Bank, 2021).

En este escenario, la prevención de la deserción escolar se convierte en un desafío prioritario para los sistemas educativos, ya que las estrategias tradicionales de seguimiento suelen ser insuficientes cuando se aplican de manera tardía o reactiva. Muchas instituciones identifican el riesgo de abandono cuando el estudiante ya presenta ausentismo prolongado, bajo rendimiento acumulado o desvinculación parcial del proceso educativo. Esta situación evidencia la necesidad de contar con herramientas que permitan anticipar el riesgo de deserción antes de que el abandono se materialice.

Frente a esta necesidad, los modelos predictivos y las técnicas de inteligencia artificial han adquirido creciente importancia en el campo educativo. La minería de datos educativos y el machine learning permiten analizar grandes volúmenes de información para identificar patrones, relaciones y tendencias asociadas al desempeño académico, la permanencia escolar y el riesgo de abandono (Romero & Ventura, 2020). A diferencia de los enfoques descriptivos tradicionales, los modelos predictivos ofrecen la posibilidad de estimar probabilidades de riesgo y apoyar la toma de decisiones basada en evidencia.

Entre los modelos más utilizados en la literatura se encuentra la regresión logística, debido a su utilidad para analizar variables categóricas y estimar la probabilidad de ocurrencia de eventos educativos, como aprobar, reprobar o abandonar los estudios. Asimismo, algoritmos como árboles de decisión, Random Forest, Support Vector Machine y redes neuronales han sido empleados en investigaciones orientadas a clasificar estudiantes según niveles de riesgo académico o probabilidad de deserción (Márquez-Vera et al., 2016; Berens et al., 2019). Sin embargo, la variedad de modelos utilizados, los distintos contextos de aplicación y la diversidad de variables analizadas dificultan establecer consensos claros sobre cuáles técnicas ofrecen mejores resultados en la predicción del abandono escolar.

A pesar del crecimiento de investigaciones sobre inteligencia artificial aplicada a la educación, persisten vacíos importantes en la literatura científica. En primer lugar, muchos estudios se concentran en contextos de educación superior, mientras que existe menor sistematización sobre su aplicación en educación básica y media. En segundo lugar, las investigaciones presentan diferencias metodológicas significativas en cuanto a bases de datos utilizadas, criterios de evaluación, métricas de desempeño y selección de variables predictoras. En tercer lugar, no siempre se discuten de manera suficiente las implicaciones éticas del uso de modelos algorítmicos en contextos educativos, especialmente en relación con la privacidad de los datos, el sesgo algorítmico y la interpretación pedagógica de los resultados (UNESCO, 2021).

En América Latina aún se requiere mayor producción científica que analice de manera integrada el uso de modelos predictivos para prevenir la deserción escolar desde una perspectiva contextualizada. Las condiciones educativas de la región, marcadas por desigualdad social, brechas digitales, diferencias urbano-rurales y limitaciones institucionales, exigen revisar críticamente qué modelos han sido aplicados, qué variables han demostrado mayor relevancia y qué limitaciones presentan los estudios existentes. Sin esta sistematización, resulta difícil orientar futuras investigaciones y diseñar estrategias educativas basadas en evidencia científica.

Por tanto, el problema central que aborda esta investigación radica en la dispersión teórica, metodológica y empírica existente en la literatura científica sobre el uso de modelos predictivos y técnicas de machine learning para la predicción de la deserción escolar. Aunque existen estudios que demuestran el potencial de estas herramientas, aún se requiere una revisión bibliográfica que permita ordenar, comparar y analizar críticamente los principales aportes científicos, modelos utilizados, variables predictoras, tendencias metodológicas y limitaciones reportadas.

En consecuencia, surge la siguiente pregunta de investigación:

¿Cuáles son los principales aportes, modelos predictivos, variables y tendencias metodológicas identificadas en la literatura científica sobre el uso de machine learning y regresión logística para la predicción de la deserción escolar?

A partir de esta pregunta, el presente artículo de revisión bibliográfica busca contribuir a la comprensión del estado actual del conocimiento sobre la aplicación de modelos predictivos en la prevención del abandono escolar, proporcionando una base teórica y analítica para futuras investigaciones, intervenciones educativas y políticas públicas orientadas a fortalecer la permanencia estudiantil.

1.3 Objetivo general

Analizar la literatura científica relacionada con la aplicación de modelos predictivos y técnicas de machine learning en la predicción de la deserción escolar, mediante una revisión bibliográfica de estudios publicados en bases de datos académicas indexadas.

2. METODOLOGÍA

2.1 Enfoque y diseño de la revisión

La presente investigación se desarrolló bajo un enfoque cualitativo-documental, mediante una revisión bibliográfica narrativa con orientación sistemática. Este enfoque permitió analizar, organizar e interpretar la producción científica relacionada con el uso de modelos predictivos, regresión logística y técnicas de *machine learning* en la predicción de la deserción escolar. Al tratarse de un artículo de revisión, no se recolectaron datos primarios ni se aplicaron instrumentos a participantes, sino que se trabajó con fuentes secundarias previamente publicadas en bases académicas y científicas.

La revisión bibliográfica fue seleccionada porque permite sintetizar conocimientos dispersos, identificar tendencias investigativas, reconocer vacíos teóricos y comparar los principales aportes de investigaciones previas sobre un fenómeno determinado (Snyder, 2019). En este sentido, el estudio

no busca comprobar hipótesis mediante datos empíricos propios, sino analizar críticamente la literatura existente sobre inteligencia artificial, analítica educativa y abandono escolar.

El diseño de la investigación fue documental, descriptivo y analítico. Fue documental porque la unidad de análisis estuvo conformada por artículos científicos, revisiones, capítulos académicos e informes especializados; descriptivo porque permitió caracterizar los modelos predictivos, variables y enfoques metodológicos presentes en la literatura; y analítico porque se orientó a interpretar los aportes, limitaciones y tendencias identificadas en los estudios revisados.

2.2 Fuentes de información y estrategia de búsqueda

La búsqueda bibliográfica se realizó en bases de datos académicas y repositorios científicos reconocidos por su relevancia en los campos de educación, ciencias sociales, inteligencia artificial, minería de datos y analítica del aprendizaje. Las principales fuentes consultadas fueron Scopus, Web of Science, ERIC, SciELO, Redalyc, Dialnet, ScienceDirect, SpringerLink y Google Scholar.

El periodo de revisión se delimitó preferentemente entre los años 2020 y 2025, con el fin de priorizar literatura reciente y actualizada. No obstante, se incorporaron algunas fuentes clásicas cuando resultaron necesarias para fundamentar conceptos centrales sobre deserción escolar, permanencia estudiantil, modelos predictivos y metodología de revisión. Esta decisión permitió integrar aportes actuales con referentes teóricos consolidados, fortaleciendo la base académica del estudio.

2.3 Criterios de selección documental

Para garantizar la pertinencia y calidad de la información analizada, se establecieron criterios de inclusión y exclusión. Se incluyeron artículos científicos, revisiones bibliográficas, revisiones sistemáticas, capítulos académicos e informes técnicos publicados por organismos reconocidos, siempre que abordaran temas relacionados con deserción escolar, abandono educativo, rendimiento académico, riesgo de permanencia estudiantil, regresión logística, minería de datos educativa o técnicas de *machine learning* aplicadas al contexto educativo.

Por otro lado, se excluyeron publicaciones duplicadas, notas periodísticas, blogs, ensayos de opinión, documentos sin autoría identificable, investigaciones sin respaldo académico y estudios que abordaban inteligencia artificial sin relación directa con el ámbito educativo. Asimismo, se descartaron trabajos que no ofrecían información suficiente sobre el procedimiento metodológico, las variables consideradas o los resultados obtenidos.

2.4 Procedimiento de revisión y análisis de la información

El procedimiento de revisión se desarrolló en tres fases. En la primera fase, se realizó la búsqueda inicial de documentos en las bases de datos seleccionadas, revisando títulos, resúmenes y palabras clave para identificar estudios potencialmente relevantes. En la segunda fase, se aplicaron los criterios de inclusión y exclusión, eliminando documentos duplicados o alejados del objetivo de la investigación. En la tercera fase, se llevó a cabo la lectura analítica de los textos seleccionados, organizando la información en una matriz bibliográfica.

La matriz de análisis permitió clasificar los estudios según autor, año de publicación, país o contexto, objetivo, tipo de modelo predictivo utilizado, variables consideradas, principales hallazgos y limitaciones reportadas. Esta organización facilitó la comparación entre investigaciones y permitió identificar tendencias comunes dentro de la literatura científica revisada.

El análisis de la información se realizó mediante síntesis temática, agrupando los hallazgos en categorías relacionadas con los factores asociados a la deserción escolar, las variables predictoras más empleadas, los modelos estadísticos y algoritmos de *machine learning* utilizados, las métricas de desempeño reportadas y los desafíos metodológicos o éticos identificados en los estudios. Esta técnica resulta adecuada en revisiones bibliográficas porque permite integrar información procedente de diferentes fuentes y construir una interpretación crítica del estado del conocimiento (Grant & Booth, 2009).

Aunque esta investigación no corresponde a una revisión sistemática estricta ni a un metaanálisis, se tomaron como referencia algunos principios del modelo PRISMA 2020, especialmente en lo relacionado con la transparencia del proceso de búsqueda, selección y organización documental. Page et al. (2021) señalan que la claridad en el procedimiento de revisión contribuye a fortalecer la trazabilidad, confiabilidad y rigor científico de los estudios bibliográficos.

Con el propósito de representar de manera gráfica el proceso de identificación, cribado, elegibilidad e inclusión de los documentos revisados, se elaboró un diagrama de flujo basado en las orientaciones generales del modelo PRISMA 2020. Este esquema permite visualizar de forma ordenada la ruta metodológica seguida para la selección documental, desde la búsqueda inicial en bases de datos académicas hasta la inclusión final de los estudios considerados pertinentes para la revisión bibliográfica.

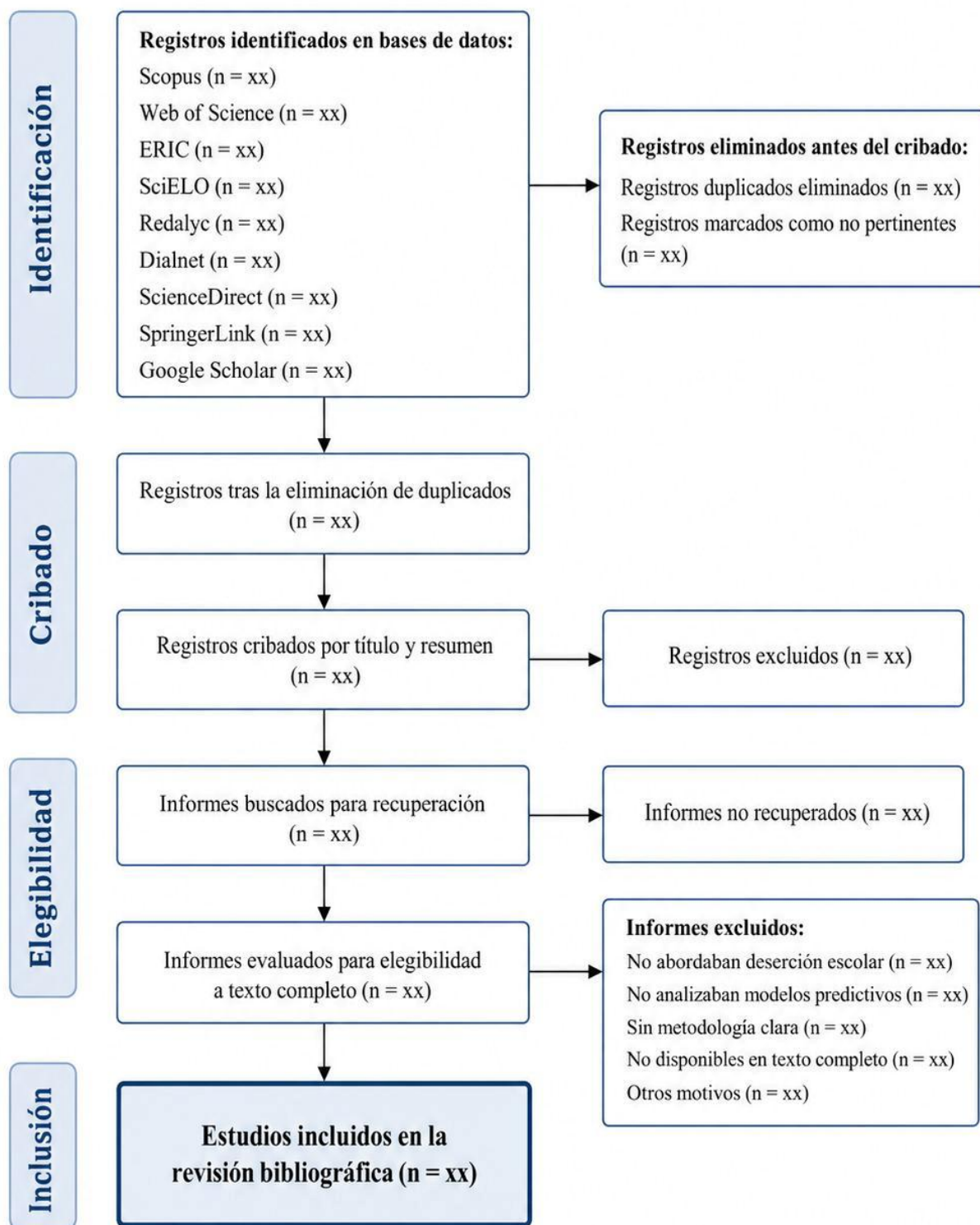


Figura 1 Diagrama de flujo PRISMA del proceso de selección de estudios

Como se observa en la Figura 1, el proceso de selección documental se estructuró en cuatro fases principales: identificación, cribado, elegibilidad e inclusión. En la fase de identificación se recopilaron registros procedentes de bases de datos académicas y repositorios científicos. Posteriormente, durante el cribado, se eliminaron documentos duplicados y estudios no pertinentes según el título y resumen. En la fase de elegibilidad se revisaron los textos completos disponibles,

excluyendo aquellos que no abordaban directamente la deserción escolar, los modelos predictivos o las técnicas de *machine learning*. Finalmente, se incluyeron los estudios que cumplieron con los criterios de pertinencia, actualidad, rigor metodológico y relación directa con el objetivo de la revisión.

3. MARCO TEÓRICO

3.1 Deserción escolar y permanencia educativa

La deserción escolar constituye uno de los fenómenos más complejos de los sistemas educativos, debido a que no responde a una causa única, sino a la interacción de múltiples factores académicos, familiares, socioeconómicos, institucionales y contextuales. En términos generales, se entiende como la interrupción de la trayectoria educativa antes de culminar un determinado nivel de formación, lo que afecta no solo el desarrollo académico del estudiante, sino también sus oportunidades futuras de inserción laboral, movilidad social y participación ciudadana.

Desde una perspectiva teórica, la deserción escolar puede comprenderse como un proceso gradual de desvinculación del estudiante con la institución educativa. Tinto (1993) plantea que la permanencia estudiantil está relacionada con el grado de integración académica y social que el estudiante logra dentro del entorno educativo. Cuando esta integración es débil, ya sea por bajo rendimiento, falta de apoyo institucional, escasa motivación o dificultades de adaptación, aumenta la probabilidad de abandono. Aunque el modelo de Tinto fue formulado inicialmente para la educación superior, sus aportes han sido ampliamente utilizados para comprender procesos de retención y abandono en distintos niveles educativos.

De manera complementaria, Rumberger (2011) sostiene que la deserción escolar debe analizarse desde una perspectiva ecológica, considerando tanto las características individuales del estudiante como las condiciones familiares, escolares y comunitarias que influyen en su permanencia. Esta visión permite superar explicaciones reduccionistas que atribuyen el abandono únicamente a la falta de interés o esfuerzo del estudiante, reconociendo que muchas trayectorias de deserción están asociadas a desigualdades estructurales, pobreza, exclusión social y limitaciones institucionales.

La literatura científica ha identificado diversos indicadores tempranos asociados al riesgo de abandono escolar. Entre los más frecuentes se encuentran el bajo rendimiento académico, la repitencia, el ausentismo, la reprobación de asignaturas, la falta de acompañamiento familiar, la necesidad de trabajar, la baja motivación y las condiciones socioeconómicas adversas (Bowers et al., 2013; Rumberger, 2011). Estos factores no actúan de manera aislada, sino que se combinan y acumulan a lo largo del tiempo, generando procesos de vulnerabilidad educativa.

En América Latina, la deserción escolar adquiere una dimensión particular debido a la persistencia de brechas sociales, económicas, digitales y territoriales. Las poblaciones rurales, los estudiantes de bajos ingresos, los grupos históricamente excluidos y quienes enfrentan condiciones de violencia o precariedad suelen presentar mayores riesgos de interrupción educativa. La UNESCO (2021) señala que la crisis generada por la pandemia de COVID-19 profundizó las desigualdades educativas, afectando especialmente a estudiantes con menor acceso a conectividad, dispositivos tecnológicos y apoyo familiar.

3.2 Factores asociados a la deserción escolar

La deserción escolar es un fenómeno multidimensional, por lo que su análisis requiere considerar diversas categorías de factores. La literatura revisada coincide en que los modelos predictivos más robustos son aquellos que integran variables académicas, socioeconómicas, familiares, institucionales y contextuales, ya que estas dimensiones permiten comprender de manera más completa las causas del abandono educativo.

Las variables académicas suelen ser las más utilizadas en los estudios predictivos, debido a que ofrecen información directa sobre la trayectoria escolar del estudiante. Indicadores como calificaciones, promedio general, asistencia, repitencia y reprobación permiten identificar señales tempranas de riesgo. Bowers et al. (2013) sostienen que los indicadores académicos son útiles para anticipar procesos de desvinculación, aunque advierten que deben interpretarse junto con otras variables para evitar conclusiones incompletas.

Las variables socioeconómicas también resultan determinantes. El ingreso familiar, el nivel educativo de los padres, la ocupación del hogar, la disponibilidad de recursos tecnológicos y las condiciones de vivienda pueden incidir directamente en la permanencia escolar. Rumberger (2011) señala que las condiciones familiares y comunitarias influyen de manera significativa en las trayectorias educativas, especialmente cuando los estudiantes enfrentan pobreza, inseguridad económica o responsabilidades laborales tempranas.

Los factores institucionales hacen referencia a las condiciones que ofrece la escuela para garantizar la permanencia. Entre ellos se encuentran la calidad de la infraestructura, el clima escolar, el acompañamiento pedagógico, la existencia de programas de apoyo, la orientación psicopedagógica, el transporte, la alimentación escolar y el acceso a recursos educativos. Freeman y Simonsen (2015) afirman que las intervenciones más efectivas para reducir la deserción combinan acciones académicas, conductuales, familiares e institucionales.

A continuación, se presenta una tabla de síntesis con las principales dimensiones asociadas a la deserción escolar identificadas en la literatura.

Tabla 1 Principales factores asociados a la deserción escolar

Dimensión	Variables frecuentes	Relación con la deserción escolar	Autores de referencia
Académica	Bajo rendimiento, repitencia, reprobación, ausentismo, promedio general	Permite identificar trayectorias escolares inestables y señales tempranas de riesgo	Bowers et al. (2013); Rumberger (2011)
Socioeconómica	Ingreso familiar, pobreza, empleo infantil, acceso a tecnología, condiciones de vivienda	Limita las oportunidades de permanencia y aumenta la vulnerabilidad educativa	Rumberger (2011); UNESCO (2021)
Familiar	Escolaridad de los padres, apoyo familiar, estructura del hogar, expectativas educativas	Puede actuar como factor protector o de riesgo según el nivel de acompañamiento	Rumberger (2011); Tinto (1993)
Institucional	Clima escolar, tutorías, infraestructura, orientación, programas de apoyo	Incide en la integración académica y social del estudiante	Freeman y Simonsen (2015); Tinto (1993)
Contextual	Ruralidad, violencia, migración, brecha digital, distancia al centro educativo	Condiciona el acceso, continuidad y calidad de la experiencia educativa	UNESCO (2021); UNESCO et al. (2021)

Fuente: Elaboración propia con base en Tinto (1993), Rumberger (2011), Bowers et al. (2013), Freeman y Simonsen (2015), UNESCO (2021).

Este conjunto de variables evidencia que la deserción escolar no puede explicarse únicamente desde el desempeño individual del estudiante. Por el contrario, requiere una lectura integral que considere las condiciones sociales, económicas e institucionales que configuran las oportunidades reales de permanencia. Esta perspectiva resulta clave para el uso de modelos predictivos, ya que la calidad de una predicción depende en gran medida de la pertinencia de las variables incluidas en el análisis.

3.3 Analítica educativa, minería de datos e inteligencia artificial en educación

El avance de la inteligencia artificial y el crecimiento de los sistemas de información educativa han transformado la manera en que las instituciones analizan los procesos de aprendizaje, rendimiento y permanencia. En este contexto, la analítica del aprendizaje, la minería de datos educativa y la analítica académica han adquirido un papel relevante en la identificación temprana de estudiantes en riesgo.

La analítica del aprendizaje se define como la medición, recopilación, análisis y presentación de datos sobre los estudiantes y sus contextos, con el propósito de comprender y optimizar el

aprendizaje y los entornos donde este ocurre (Siemens y Long, 2011). Esta definición destaca que el uso de datos en educación no debe limitarse a fines administrativos, sino orientarse a mejorar los procesos pedagógicos, fortalecer el acompañamiento estudiantil y apoyar la toma de decisiones.

La minería de datos educativa, por su parte, se enfoca en el uso de métodos estadísticos, computacionales y algorítmicos para descubrir patrones en grandes conjuntos de datos educativos. Romero y Ventura (2020) señalan que este campo ha evolucionado significativamente, permitiendo desarrollar aplicaciones para predecir rendimiento académico, detectar estudiantes en riesgo, personalizar procesos de aprendizaje y evaluar estrategias pedagógicas. En el caso de la deserción escolar, su utilidad radica en la posibilidad de identificar combinaciones de variables que anticipan el abandono.

La inteligencia artificial aplicada a la educación ofrece herramientas para procesar información académica, socioeconómica e institucional con el fin de generar modelos de apoyo a la decisión. Sin embargo, su implementación requiere criterios pedagógicos y éticos. La UNESCO (2021) advierte que el uso de inteligencia artificial en educación debe promover la inclusión, la equidad, la transparencia y la protección de los derechos de los estudiantes. Por ello, los sistemas predictivos deben entenderse como instrumentos de apoyo, no como mecanismos automáticos de clasificación o exclusión.

En la prevención de la deserción escolar, estas herramientas permiten transitar de un modelo reactivo hacia un modelo preventivo. Tradicionalmente, las instituciones suelen identificar el abandono cuando el estudiante ya ha acumulado inasistencias, bajo rendimiento o desvinculación evidente. En cambio, los sistemas basados en datos pueden detectar señales tempranas y generar alertas para que docentes, orientadores y directivos intervengan oportunamente.

3.4 Modelos predictivos aplicados a la deserción escolar

Los modelos predictivos son herramientas estadísticas y computacionales que permiten estimar la probabilidad de ocurrencia de un evento futuro a partir de datos previos. En el ámbito educativo, se han utilizado para anticipar bajo rendimiento académico, reprobación, ausentismo, abandono escolar y necesidades de apoyo pedagógico. Su principal aporte consiste en transformar datos institucionales en información útil para la toma de decisiones.

Uno de los modelos más empleados en la literatura es la regresión logística. Este método permite estimar la probabilidad de que ocurra un evento dicotómico, como abandonar o no abandonar los estudios, aprobar o reprobado, estar o no estar en riesgo académico. Hosmer et al. (2013) explican que la regresión logística es adecuada cuando se busca analizar la relación entre una variable

dependiente binaria y un conjunto de variables predictoras. Su principal ventaja en educación es la interpretabilidad, ya que permite comprender qué factores aumentan o reducen la probabilidad de deserción.

Además de la regresión logística, los algoritmos de aprendizaje automático han ampliado las posibilidades de predicción educativa. Técnicas como árboles de decisión, Random Forest, máquinas de soporte vectorial, redes neuronales artificiales, Naive Bayes y métodos de ensamble han sido utilizadas para clasificar estudiantes según niveles de riesgo. James et al. (2021) señalan que los métodos de aprendizaje automático resultan especialmente útiles cuando existen relaciones complejas o no lineales entre variables, así como grandes volúmenes de datos.

Márquez-Vera et al. (2016) aplicaron técnicas de minería de datos para predecir tempranamente la deserción en estudiantes de secundaria, demostrando que los modelos predictivos pueden identificar patrones de riesgo antes de que ocurra el abandono definitivo. De manera similar, Berens et al. (2019) utilizaron datos administrativos y aprendizaje automático para detectar estudiantes universitarios en riesgo de deserción, destacando el valor de la información institucional para orientar intervenciones tempranas.

Tabla 2 Modelos predictivos utilizados en estudios sobre deserción escolar

Modelo	Características principales	Ventajas en educación	Limitaciones
Regresión logística	Estima la probabilidad de ocurrencia de un evento binario	Alta interpretabilidad; permite identificar peso de variables predictoras	Puede ser limitada ante relaciones no lineales complejas
Árboles de decisión	Clasifican casos mediante reglas jerárquicas	Fácil comprensión; útil para generar reglas de intervención	Puede presentar sobreajuste si no se controla adecuadamente
Random Forest	Combina múltiples árboles de decisión	Alta precisión; reduce el riesgo de sobreajuste	Menor interpretabilidad que un árbol simple
Support Vector Machine	Clasifica datos mediante hiperplanos de separación	Buen desempeño en problemas complejos de clasificación	Difícil interpretación para usuarios no técnicos
Redes neuronales artificiales	Modelan relaciones complejas entre variables	Alta capacidad predictiva en grandes volúmenes de datos	Baja interpretabilidad; requiere mayor capacidad técnica
Naive Bayes	Modelo probabilístico basado en el teorema de Bayes	Simple, rápido y útil con bases grandes	Supone independencia entre variables, lo cual no siempre se cumple

Fuente: Elaboración propia con base en Hosmer et al. (2013), Márquez-Vera et al. (2016), Berens et al. (2019), Romero y Ventura (2020) y James et al. (2021).

Esta tabla evidencia que no existe un único modelo ideal para la predicción de la deserción escolar. La selección depende del objetivo del estudio, el tipo de datos disponibles, el nivel educativo

analizado, la necesidad de interpretación y el contexto institucional donde se implementará el sistema predictivo. Por esta razón, la literatura recomienda comparar distintos modelos y evaluar su desempeño mediante métricas apropiadas.

3.5 Evaluación del desempeño predictivo

La evaluación de los modelos predictivos es un componente fundamental en los estudios sobre deserción escolar, ya que permite determinar si el modelo clasifica adecuadamente a los estudiantes en riesgo. Entre las métricas más utilizadas se encuentran la exactitud, la precisión, la sensibilidad, la especificidad, el valor F1, la matriz de confusión y el área bajo la curva ROC. Estas métricas permiten analizar el desempeño del modelo desde diferentes perspectivas.

La exactitud indica el porcentaje total de clasificaciones correctas, pero puede ser insuficiente cuando existe desbalance entre estudiantes que abandonan y quienes permanecen. La sensibilidad permite identificar la capacidad del modelo para detectar correctamente a los estudiantes que efectivamente están en riesgo, mientras que la especificidad mide la capacidad para reconocer a quienes no presentan riesgo. En contextos educativos, la sensibilidad suele ser especialmente importante, porque no detectar a un estudiante vulnerable puede impedir una intervención oportuna.

La curva ROC y el área bajo la curva permiten evaluar la capacidad discriminativa del modelo. Un valor AUC alto indica que el modelo diferencia adecuadamente entre estudiantes con riesgo y sin riesgo. Sin embargo, estas métricas deben interpretarse con cuidado. Como señalan James et al. (2021), un buen desempeño estadístico no garantiza por sí solo utilidad práctica, especialmente si el modelo no es interpretable o si sus resultados no se traducen en acciones institucionales concretas.

En educación la evaluación de un modelo predictivo debe considerar tanto su rendimiento técnico como su utilidad pedagógica. No basta con identificar estudiantes en riesgo; es necesario que la información generada permita activar procesos de tutoría, apoyo académico, orientación familiar, acompañamiento psicosocial o fortalecimiento institucional. En este sentido, los modelos predictivos deben formar parte de sistemas integrales de prevención y no operar como herramientas aisladas.

3.6 Consideraciones éticas en el uso de modelos predictivos

El uso de modelos predictivos en educación plantea importantes desafíos éticos relacionados con la privacidad de los datos, la transparencia algorítmica, el consentimiento informado, la equidad y el riesgo de estigmatización. Slade y Prinsloo (2013) sostienen que la analítica del aprendizaje debe desarrollarse bajo principios de responsabilidad institucional, uso justo de la información, protección de la privacidad y claridad sobre los fines del análisis.

Uno de los principales riesgos es el etiquetamiento de estudiantes. Cuando un modelo clasifica a un estudiante como de “alto riesgo”, existe la posibilidad de que docentes o instituciones generen expectativas negativas sobre su desempeño. Por ello, la clasificación predictiva no debe utilizarse como diagnóstico definitivo, sino como una señal de alerta que permita activar apoyos oportunos. El objetivo debe ser acompañar, no sancionar ni excluir.

Otro desafío relevante es el sesgo algorítmico. Si los datos históricos utilizados para entrenar un modelo reflejan desigualdades sociales, económicas, territoriales o étnicas, el algoritmo puede reproducir o amplificar dichas desigualdades. La UNESCO (2021) advierte que la inteligencia artificial en educación debe estar orientada por principios de justicia, inclusión y no discriminación. Esto implica revisar críticamente las variables utilizadas, los resultados generados y las decisiones institucionales derivadas del modelo.

Los modelos predictivos, la regresión logística y el aprendizaje automático ofrecen un alto potencial para prevenir la deserción escolar, siempre que sean utilizados desde una perspectiva pedagógica, contextualizada y ética. Estas herramientas pueden contribuir a la toma de decisiones basada en evidencia, pero no reemplazan el juicio docente ni la comprensión integral de las trayectorias estudiantiles. Su verdadero valor radica en apoyar intervenciones tempranas que fortalezcan la permanencia, la equidad y la calidad educativa.

4. Resultados

Los resultados de la revisión bibliográfica evidencian que la predicción de la deserción escolar se ha consolidado como una línea de investigación relevante dentro de la analítica educativa, especialmente por el avance de la inteligencia artificial, la minería de datos y los algoritmos de *machine learning*. La literatura reciente muestra que los estudios ya no se enfocan únicamente en describir las causas del abandono escolar, sino en anticipar el riesgo mediante modelos predictivos capaces de identificar patrones asociados a bajo rendimiento, ausentismo, repitencia, condiciones socioeconómicas, acceso tecnológico y participación académica (Romero & Ventura, 2020; Psyridou et al., 2024).

Una primera tendencia identificada es el tránsito desde enfoques estadísticos tradicionales hacia modelos computacionales más complejos. La regresión logística continúa siendo ampliamente utilizada por su capacidad interpretativa, especialmente cuando se busca estimar la probabilidad de abandono a partir de variables académicas y socioeconómicas. No obstante, los estudios recientes muestran una presencia creciente de algoritmos como árboles de decisión, Random Forest, redes neuronales artificiales, Support Vector Machine, XGBoost y CatBoost, debido a su capacidad para

trabajar con bases de datos amplias y detectar relaciones no lineales entre variables (Jiménez-Gutiérrez et al., 2024; Romero & Liao, 2025).

En relación con los modelos más empleados, la evidencia indica que Random Forest y los árboles de decisión aparecen de manera frecuente en investigaciones sobre predicción de abandono escolar, debido a su equilibrio entre desempeño predictivo e interpretabilidad. Asimismo, modelos como XGBoost y redes neuronales han mostrado altos niveles de precisión en contextos donde existen bases de datos estructuradas y de gran tamaño. Romero y Liao (2025), por ejemplo, compararon modelos estadísticos y de aprendizaje automático para predecir la deserción universitaria, encontrando que XGBoost obtuvo el mejor desempeño según la métrica F1, mientras que los modelos generalizados ofrecieron ventajas en términos de interpretación.

También se identificó que los estudios más recientes incorporan técnicas de inteligencia artificial explicable, como SHAP y LIME, con el propósito de mejorar la transparencia de los modelos predictivos. Este aspecto resulta fundamental en educación, ya que no basta con clasificar a un estudiante como “en riesgo”, sino que es necesario comprender qué variables explican dicha clasificación. Khatun et al. (2025) integraron regresión logística, Random Forest, XGBoost y herramientas de inteligencia artificial explicable para predecir el abandono escolar, evidenciando que la combinación de modelos predictivos con técnicas interpretativas permite generar información más útil para la toma de decisiones educativas.

Respecto a las variables predictoras, los resultados de la revisión muestran que los indicadores académicos siguen siendo los más utilizados. Entre ellos destacan las calificaciones, el promedio general, la asistencia, la repitencia, la reprobación de asignaturas, los créditos aprobados y el rendimiento en periodos iniciales de formación. Romero y Liao (2025) identificaron que las calificaciones del segundo semestre y el número de unidades aprobadas fueron variables altamente relevantes para predecir la deserción universitaria. Esto confirma que el desempeño académico temprano constituye una señal importante para anticipar trayectorias de riesgo.

No obstante, la literatura actual también destaca la importancia de integrar variables socioeconómicas, familiares y contextuales. Factores como el nivel de ingresos, la escolaridad de los padres, el acceso a recursos tecnológicos, la condición de becario, el tipo de institución, la edad, el género y el lugar de residencia aparecen con frecuencia en los estudios revisados. Khatun et al. (2025) encontraron que variables como edad, sexo, nivel educativo alcanzado, índice de riqueza y educación de los padres resultaron relevantes para explicar el abandono escolar. Estos hallazgos refuerzan la idea de que la deserción escolar no puede analizarse únicamente desde el rendimiento académico, sino desde una perspectiva multidimensional.

Otro resultado relevante es la incorporación de variables conductuales y de interacción digital. En estudios basados en plataformas virtuales de aprendizaje, los registros de actividad en entornos como Moodle han sido utilizados para anticipar abandono o fracaso académico. Marcolino et al. (2025) demostraron que los datos de interacción en plataformas educativas pueden aportar señales tempranas sobre el riesgo de deserción, especialmente cuando se analizan patrones de participación, frecuencia de acceso y actividad semanal. Este hallazgo resulta importante en contextos donde la educación virtual o híbrida genera grandes volúmenes de datos sobre el comportamiento estudiantil.

La revisión también permitió identificar que el momento de recolección de datos influye significativamente en la capacidad predictiva de los modelos. Psyridou et al. (2024) demostraron que es posible predecir el abandono en educación secundaria superior utilizando datos longitudinales recolectados desde etapas tempranas de la escolaridad. Su estudio evidenció que los modelos pueden anticipar el riesgo desde el final de la educación primaria, aunque con desempeños moderados, lo cual sugiere que el abandono escolar se construye progresivamente y no aparece de forma repentina.

En cuanto a las métricas de evaluación, los estudios revisados utilizan con frecuencia exactitud, precisión, sensibilidad, especificidad, F1-score, matriz de confusión y área bajo la curva ROC. Sin embargo, la literatura advierte que la exactitud no siempre es suficiente para valorar un modelo de predicción de deserción, especialmente cuando las bases de datos están desbalanceadas. En estos casos, métricas como sensibilidad, recall y F1-score adquieren mayor relevancia, ya que permiten evaluar la capacidad del modelo para identificar correctamente a los estudiantes que realmente se encuentran en riesgo (Romero & Liao, 2025; Marcolino et al., 2025).

Un hallazgo transversal en la literatura es que los modelos predictivos más útiles no son necesariamente aquellos con mayor precisión estadística, sino aquellos que combinan buen desempeño, interpretabilidad y aplicabilidad institucional. En educación, un modelo altamente complejo puede resultar limitado si sus resultados no son comprensibles para docentes, orientadores o directivos. Por ello, la literatura reciente insiste en la necesidad de desarrollar sistemas predictivos explicables, contextualizados y orientados a la intervención temprana (Khatun et al., 2025; Marcolino et al., 2025).

De manera general, los resultados muestran que la predicción de la deserción escolar ha avanzado hacia enfoques más integrales, en los que se combinan datos académicos, socioeconómicos, familiares, conductuales y tecnológicos. Esta integración permite construir modelos más sensibles a la realidad educativa de los estudiantes y evita reducir el abandono escolar únicamente a problemas de rendimiento. En este sentido, la revisión evidencia que los modelos predictivos pueden contribuir

significativamente a la prevención del abandono escolar, siempre que sean utilizados como herramientas de apoyo pedagógico y no como mecanismos de clasificación rígida.

A continuación, se sintetizan los principales hallazgos identificados en la literatura revisada.

Tabla 3 Principales hallazgos sobre modelos predictivos aplicados a la deserción escolar

Categoría analizada	Hallazgos principales	Referencias
Modelos más utilizados	Regresión logística, árboles de decisión, Random Forest, redes neuronales, SVM, XGBoost y CatBoost	Jiménez-Gutiérrez et al. (2024); Romero y Liao (2025); Marcolino et al. (2025)
Variables académicas	Calificaciones, promedio, asistencia, repitencia, créditos aprobados y reprobación son predictores frecuentes	Romero y Liao (2025); Psyridou et al. (2024)
Variables socioeconómicas	Ingreso familiar, escolaridad de padres, edad, género, becas y condiciones del hogar influyen en el riesgo de abandono	Khatun et al. (2025); Romero y Liao (2025)
Datos longitudinales	El riesgo de abandono puede anticiparse desde etapas tempranas mediante seguimiento longitudinal	Psyridou et al. (2024)
Datos digitales	Los registros de plataformas virtuales permiten identificar patrones de participación asociados al abandono	Marcolino et al. (2025)
Explicabilidad	SHAP y LIME fortalecen la interpretación de modelos complejos y facilitan la toma de decisiones educativas	Khatun et al. (2025)
Métricas de evaluación	F1-score, recall, sensibilidad y AUC son relevantes cuando existen datos desbalanceados	Romero y Liao (2025); Marcolino et al. (2025)

Fuente: Elaboración propia con base en Psyridou et al. (2024), Jiménez-Gutiérrez et al. (2024), Romero y Liao (2025), Khatun et al. (2025) y Marcolino et al. (2025).

En síntesis, la revisión bibliográfica muestra que los modelos predictivos y las técnicas de *machine learning* representan herramientas relevantes para anticipar la deserción escolar y fortalecer la toma de decisiones basada en evidencia. Sin embargo, su implementación debe considerar no solo el rendimiento técnico del modelo, sino también la calidad de los datos, la interpretación pedagógica de los resultados, la equidad en el tratamiento de los estudiantes y la posibilidad de activar intervenciones oportunas. De este modo, la inteligencia artificial aplicada a la educación puede contribuir a la permanencia escolar siempre que se utilice desde una perspectiva ética, contextualizada y orientada al acompañamiento estudiantil.

5. DISCUSIÓN

Los resultados de la revisión bibliográfica evidencian que los modelos predictivos y las técnicas de *machine learning* se han convertido en herramientas relevantes para anticipar la deserción

escolar y fortalecer la toma de decisiones educativas. La literatura reciente muestra que el abandono escolar no ocurre de manera repentina, sino que responde a un proceso progresivo influido por factores académicos, socioeconómicos, familiares, institucionales y contextuales (Psyridou et al., 2024; Romero & Ventura, 2020).

Uno de los hallazgos más importantes es que la regresión logística continúa siendo un modelo ampliamente utilizado por su capacidad interpretativa. Este método permite estimar la probabilidad de abandono escolar y, al mismo tiempo, identificar qué variables tienen mayor influencia en el riesgo de deserción. Sin embargo, los estudios recientes también muestran un crecimiento en el uso de algoritmos como Random Forest, XGBoost, redes neuronales y máquinas de soporte vectorial, debido a su mayor capacidad para procesar grandes volúmenes de datos y detectar relaciones complejas entre variables (Jiménez-Gutiérrez et al., 2024; Romero & Liao, 2025).

No obstante, la elección del modelo predictivo no debe depender únicamente de su precisión estadística. En el contexto educativo, también resulta fundamental que los resultados sean comprensibles para docentes, directivos y responsables de políticas públicas. Por ello, la literatura actual destaca la importancia de la inteligencia artificial explicable, ya que permite conocer las razones por las cuales un estudiante es clasificado en riesgo y facilita el diseño de intervenciones oportunas (Khatun et al., 2025).

Otro aspecto relevante es que las variables académicas, como calificaciones, asistencia, repitencia y rendimiento previo, siguen siendo los predictores más frecuentes en los estudios revisados. Sin embargo, estos indicadores no explican por sí solos la deserción escolar. La incorporación de variables socioeconómicas, familiares, digitales y contextuales permite construir modelos más integrales y cercanos a la realidad de los estudiantes (Matz et al., 2023; Psyridou et al., 2024).

Desde esta perspectiva, los modelos predictivos deben entenderse como herramientas de apoyo y no como mecanismos automáticos de clasificación. Su finalidad no debe ser etiquetar a los estudiantes, sino identificar señales tempranas de riesgo para activar procesos de acompañamiento académico, orientación psicosocial, apoyo familiar y estrategias institucionales de permanencia. Slade y Prinsloo (2013) advierten que el uso de analítica educativa debe considerar principios de privacidad, transparencia, responsabilidad y justicia.

En síntesis, la revisión permite afirmar que la aplicación de regresión logística, minería de datos y *machine learning* posee un alto potencial para prevenir la deserción escolar. Sin embargo, su efectividad depende de la calidad de los datos, la correcta selección de variables, la interpretación

pedagógica de los resultados y el uso ético de la información. Por tanto, la inteligencia artificial aplicada a la educación debe orientarse hacia la prevención, la inclusión y el fortalecimiento de trayectorias escolares, evitando prácticas de discriminación, estigmatización o exclusión (UNESCO, 2021).

6. CONCLUSIONES

La revisión bibliográfica permitió concluir que la deserción escolar es un fenómeno multidimensional que no puede explicarse únicamente desde el rendimiento académico del estudiante. La literatura científica evidencia que el abandono educativo está asociado a la interacción de factores académicos, socioeconómicos, familiares, institucionales, digitales y contextuales, por lo que su prevención requiere enfoques integrales y estrategias de detección temprana.

Se concluye que los modelos predictivos y las técnicas de *machine learning* constituyen herramientas relevantes para anticipar el riesgo de deserción escolar. Su principal aporte radica en la posibilidad de identificar patrones de vulnerabilidad antes de que el abandono se materialice, permitiendo a las instituciones educativas diseñar intervenciones oportunas, focalizadas y basadas en evidencia.

La regresión logística continúa siendo uno de los modelos más utilizados en la literatura debido a su capacidad interpretativa y a su utilidad para estimar probabilidades de riesgo. Sin embargo, los estudios recientes muestran un crecimiento importante en el uso de algoritmos como Random Forest, árboles de decisión, XGBoost, redes neuronales artificiales y máquinas de soporte vectorial, los cuales ofrecen mayor capacidad para analizar grandes volúmenes de datos y relaciones complejas entre variables.

Los hallazgos también permiten concluir que las variables académicas, como calificaciones, asistencia, repitencia, reprobación y rendimiento previo, son los predictores más frecuentes en los estudios revisados. No obstante, los modelos más completos son aquellos que incorporan además variables socioeconómicas, familiares, digitales y contextuales, ya que permiten comprender la deserción escolar desde una perspectiva más amplia y cercana a la realidad de los estudiantes.

Se identificó que la precisión estadística no debe ser el único criterio para valorar la utilidad de un modelo predictivo. En el ámbito educativo, resulta fundamental que los modelos sean interpretables, explicables y aplicables a la toma de decisiones pedagógicas. Por ello, el uso de herramientas de inteligencia artificial explicable representa una tendencia relevante para fortalecer la transparencia y la comprensión de los resultados.

Se concluye que la aplicación de modelos predictivos en educación debe desarrollarse bajo criterios éticos, evitando el etiquetamiento, la discriminación o la exclusión de estudiantes clasificados como “en riesgo”. Estas herramientas deben utilizarse como apoyo para el acompañamiento académico, la orientación psicosocial, la intervención institucional y el fortalecimiento de la permanencia escolar. En este sentido, la inteligencia artificial aplicada a la educación tiene un alto potencial para contribuir a sistemas educativos más preventivos, inclusivos y basados en evidencia científica.

FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación.

CONFLICTO DE INTERESES

Los Autores declaran que no existe conflicto de intereses con su investigación

CONTRIBUCIÓN DE AUTORÍA

En concordancia con la taxonomía establecida internacionalmente para la asignación de créditos a autores de artículos científicos (<https://credit.niso.org/>). Los autores declaran sus contribuciones en la siguiente matriz:

<i>Participar activamente en:</i>	<i>Autor 1</i>	<i>Autor 2</i>	<i>Autor 3</i>	<i>Autor 4</i>
<i>Conceptualización</i>	X	X	X	X
<i>Análisis formal</i>	X	X	X	X
<i>Adquisición de fondos</i>	X			
<i>Investigación</i>	X	X	X	X
<i>Metodología</i>	X	X	X	X
<i>Administración del proyecto</i>		X	X	X
<i>Recursos</i>	X	X	X	X
<i>Redacción –borrador original</i>	X	X	X	X
<i>Redacción –revisión y edición</i>	X	X	X	X
<i>La discusión de los resultados</i>	X	X	X	X
<i>Revisión y aprobación de la versión final del trabajo.</i>	X	X	X	X

8. REFERENCIAS BIBLIOGRÁFICAS:

- Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk: Predicting student dropouts using administrative student data from German universities and machine learning methods. *Journal of Educational Data Mining*, 11(3), 1–41.
- Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, 96(2), 77–100. <https://doi.org/10.1353/hsj.2013.0000>
- Bustamante Neira, G. J., & Cabrera Berrezueta, L. B. (2022). Factores que inciden en el rendimiento académico de los estudiantes de bachillerato en el cantón Sucúa-Ecuador. *Ciencia Digital*, 6(4), 97–115. <https://doi.org/10.33262/cienciadigital.v6i4.2338>

- Comisión Económica para América Latina y el Caribe. (2023). Panorama Social de América Latina y el Caribe 2023: La inclusión laboral como eje central para el desarrollo social inclusivo. Naciones Unidas.
- Delgado Aguilar, R. C., Gutiérrez Reynoso, M. R., & Dueñas Bejar, J. (2014). Predicción del rendimiento académico en la asignatura de Matemática Básica en la UNALM utilizando la técnica de regresión logística binaria. *Anales Científicos*, 75(2), 310–315.
- Freeman, J., & Simonsen, B. (2015). Examining the impact of policy and practice interventions on high school dropout and school completion rates: A systematic review of the literature. *Review of Educational Research*, 85(2), 205–248. <https://doi.org/10.3102/0034654314554431>
- González-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., & García-Castelán, R. M. G. (2023). Predictive analytics study to determine undergraduate students at risk of dropout. *Frontiers in Education*, 8, 1244686. <https://doi.org/10.3389/feduc.2023.1244686>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Hernández-Sampieri, R., & Mendoza, C. P. (2021). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (2.ª ed.). McGraw-Hill.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- 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, 3957. <https://doi.org/10.1038/s41598-024-53576-1>
- Khatun, M. R., Mim, M. A., Tasin, M. M., & Hossain, M. M. (2025). A hybrid framework of statistical, machine learning, and explainable AI methods for school dropout prediction. *PLOS ONE*, 20(9), e0331917. <https://doi.org/10.1371/journal.pone.0331917>
- Krüger, J. G. C., Britto, A. S., Jr., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications*, 233, 120933. <https://doi.org/10.1016/j.eswa.2023.120933>
- Marcolino, M. R., Porto, T. R., Primo, T. T., Targino, R., Ramos, V., Queiroga, E. M., Munoz, R., & Cechinel, C. (2025). Student dropout prediction through machine learning optimization: Insights from Moodle log data. *Scientific Reports*, 15, 9840. <https://doi.org/10.1038/s41598-025-93918-1>
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Fardoun, H. M., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, 33(1), 107–124. <https://doi.org/10.1111/exsy.12135>
- 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, 5705. <https://doi.org/10.1038/s41598-023-32484-w>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

- Psyridou, M., Prezja, F., Torppa, M., Lerkkanen, M.-K., Poikkeus, A.-M., & Vasalampi, K. (2024). Machine learning predicts upper secondary education dropout as early as the end of primary school. *Scientific Reports*, 14, 12956. <https://doi.org/10.1038/s41598-024-63629-0>
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- Romero, S., & Liao, X. (2025). Statistical and machine learning models for predicting university dropout and scholarship impact. *PLOS ONE*, 20(6), e0325047. <https://doi.org/10.1371/journal.pone.0325047>
- Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Harvard University Press.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press.
- UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO. <https://doi.org/10.54675/PCSP7350>
- UNESCO. (2024). *Global education monitoring report 2024/5: Leadership in education: Lead for learning*. UNESCO.
- UNESCO, UNICEF, & World Bank. (2021). *The state of the global education crisis: A path to recovery*. UNESCO, UNICEF, & World Bank.
- World Bank. (2021). *Acting now to protect the human capital of our children: The costs of and response to COVID-19 pandemic's impact on the education sector in Latin America and the Caribbean*. World Bank.