

Modelo predictivo basado en inteligencia artificial para identificar estudiantes en riesgo de abandono escolar en instituciones educativas
Artificial Intelligence-Based Predictive Model to Identify Students at Risk of School Dropout in Educational Institutions

MSc. Tenorio Arroyo Vilma Zulema, MSc. Tenorio Arroyo Linner Maritza, MSc. Lobato Mayanquer Andrea Viviana
MSc. Escobar Luna Monica Patricia.

**CIENCIA, TECNOLOGÍA Y
SOCIEDAD**

Julio - diciembre, V°7-N°2; 2026

Recibido: 01 -09-2026

Aceptado:04 -09-2026

Publicado:31- 12-2026

PAIS

- Esmeraldas - Ecuador
- Esmeraldas - Ecuador
- Carchi - Ecuador
- Carchi - Ecuador

INSTITUCION

- Unidad Educativa Ramón Bedoya Navia
- Unidad Educativa 21 de Septiembre
- Unidad Educativa El Angel
- Unidad Educativa Piartal

CORREO:

- ✉ vilma.tenorio@docentes.educacion.edu.ec
- ✉ linner.tenorio@docentes.educacion.edu.ec
- ✉ andrea.lobato@docentes.educacion.edu.ec
- ✉ monicap.escobar@docentes.educacion.edu.ec

ORCID:

- 🌐 <https://orcid.org/0009-0006-5642-2757>
- 🌐 <https://orcid.org/0009-0001-9394-8452>
- 🌐 <https://orcid.org/0009-0000-9029-7220>
- 🌐 <https://orcid.org/0009-0002-7683-4426>

FORMATO DE

Tenorio, V., Tenorio, L., Lobato, A. & Escobar, M. (2026). *Modelo predictivo basado en inteligencia artificial para identificar estudiantes en riesgo de abandono escolar en instituciones educativas*. Revista G-ner@ndo, V°7 (N°2), p. 3671 – 3682.

Resumen

El abandono escolar constituye una problemática educativa compleja que afecta la trayectoria académica, el desarrollo social y las oportunidades futuras de los estudiantes. La identificación temprana de factores asociados a la deserción permite generar estrategias preventivas orientadas a garantizar la permanencia estudiantil. El presente estudio tiene como objetivo desarrollar un modelo predictivo basado en inteligencia artificial para identificar estudiantes con riesgo de abandono escolar en instituciones educativas. La investigación se desarrollará bajo un enfoque cuantitativo, con diseño no experimental, alcance descriptivo-predictivo y corte transversal. Se utilizarán técnicas de minería de datos y aprendizaje automático mediante algoritmos de clasificación, considerando variables académicas, socioeconómicas, familiares y de participación escolar. La información será obtenida mediante bases de datos institucionales y registros académicos de estudiantes de educación secundaria. Para la construcción del modelo se aplicarán técnicas de procesamiento de datos, selección de variables relevantes y evaluación mediante métricas de precisión, sensibilidad y exactitud predictiva. Se espera obtener un sistema inteligente capaz de identificar patrones asociados al riesgo de abandono escolar y proporcionar información útil para la toma de decisiones educativas. Los resultados contribuirán al fortalecimiento de estrategias institucionales de acompañamiento académico, permitiendo intervenciones tempranas y personalizadas para reducir los índices de deserción escolar.

Palabras clave: inteligencia artificial; aprendizaje automático; abandono escolar; modelo predictivo; educación.

Abstract

School dropout represents a complex educational problem that affects students' academic trajectories, social development, and future opportunities. Early identification of factors associated with dropout enables the implementation of preventive strategies aimed at improving student retention. This study aims to develop an artificial intelligence-based predictive model to identify students at risk of school dropout in educational institutions. The research follows a quantitative approach with a non-experimental, descriptive-predictive and cross-sectional design. Data mining and machine learning techniques will be applied using classification algorithms based on academic, socioeconomic, family and school participation variables. Data will be collected from institutional databases and academic records of secondary education students. The predictive model will be developed through data preprocessing, feature selection and evaluation using accuracy, precision, recall and predictive performance metrics. The expected results include an intelligent system capable of detecting dropout risk patterns and providing relevant information for educational decision-making. This research contributes to strengthening institutional intervention strategies through early and personalized support actions aimed at reducing school dropout rates.

Keywords: artificial intelligence; machine learning; school dropout; predictive model; education.

Introducción

El abandono escolar representa uno de los principales desafíos que enfrentan los sistemas educativos a nivel mundial, debido a sus consecuencias en la trayectoria académica, inserción laboral y desarrollo social de los individuos. Este fenómeno no responde exclusivamente a factores académicos, sino que surge como resultado de la interacción entre condiciones económicas, familiares, personales e institucionales. En este sentido, las instituciones educativas requieren herramientas innovadoras que permitan identificar anticipadamente estudiantes con mayor probabilidad de abandonar sus estudios y establecer estrategias preventivas fundamentadas en evidencia científica (OECD, 2021).

Tradicionalmente, la identificación del riesgo de abandono escolar se ha realizado mediante análisis retrospectivos basados en indicadores como bajas calificaciones, ausentismo o repitencia académica. Sin embargo, estos métodos presentan limitaciones debido a que generalmente detectan el problema cuando la situación ya se encuentra avanzada. Frente a esta realidad, la inteligencia artificial (IA) surge como una alternativa tecnológica capaz de procesar grandes volúmenes de información educativa y descubrir patrones complejos asociados al comportamiento estudiantil (Romero & Ventura, 2024).

La incorporación de técnicas de aprendizaje automático dentro del campo educativo ha permitido desarrollar modelos predictivos capaces de anticipar eventos académicos como bajo rendimiento, fracaso escolar y abandono estudiantil. Estos modelos utilizan algoritmos computacionales que aprenden a partir de datos históricos para establecer relaciones entre variables predictoras y resultados esperados. Según Pérez Niño et al. (2024), los modelos basados en Random Forest, redes neuronales artificiales y algoritmos de clasificación presentan resultados favorables para identificar estudiantes con riesgo de deserción.

La analítica del aprendizaje constituye actualmente un área estratégica dentro de la transformación digital educativa, debido a que permite convertir los datos generados por los estudiantes en información útil para la toma de decisiones. A través del análisis de registros académicos, plataformas virtuales, asistencia y participación estudiantil, es posible construir sistemas inteligentes de alerta temprana que apoyen la intervención docente y administrativa (Ifenthaler & Yau, 2021).

Diversas investigaciones recientes evidencian que los factores académicos mantienen una relación significativa con el riesgo de abandono escolar. Variables como disminución del promedio académico, bajo cumplimiento de tareas y reducción de participación en actividades educativas han demostrado ser indicadores relevantes dentro de modelos predictivos. Los

algoritmos de inteligencia artificial permiten integrar estas variables y generar predicciones más precisas que los métodos estadísticos tradicionales (Kumar et al., 2022).

Además de los factores académicos, los modelos predictivos actuales incorporan variables sociales y contextuales debido al carácter multidimensional del abandono escolar. Aspectos relacionados con situación económica familiar, distancia geográfica, apoyo parental y condiciones emocionales influyen directamente en la permanencia educativa. Los sistemas basados en inteligencia artificial permiten analizar estas relaciones complejas y establecer perfiles de riesgo individualizados (Mihoubi et al., 2025).

En América Latina, la deserción escolar continúa siendo una problemática asociada principalmente a desigualdades sociales, limitaciones económicas y diferencias en acceso a oportunidades educativas. En Ecuador, aunque se han implementado políticas orientadas a mejorar la permanencia estudiantil, todavía existen dificultades para identificar tempranamente grupos vulnerables. Por ello, la aplicación de modelos predictivos representa una oportunidad para fortalecer los procesos institucionales de seguimiento académico.

Los avances recientes en inteligencia artificial explicable (Explainable Artificial Intelligence, XAI) han permitido superar una de las principales críticas hacia los modelos predictivos: la falta de transparencia en sus decisiones. Estas herramientas permiten conocer cuáles variables tienen mayor influencia en la predicción del abandono escolar, favoreciendo procesos de intervención más éticos y comprensibles para docentes y autoridades educativas (Alamri et al., 2025).

Los sistemas de alerta temprana basados en inteligencia artificial permiten pasar de una educación reactiva hacia una educación preventiva, donde las instituciones pueden actuar antes de que ocurra la interrupción definitiva de los estudios. La combinación entre modelos predictivos y estrategias pedagógicas personalizadas puede mejorar los índices de retención escolar y favorecer una atención educativa centrada en las necesidades individuales del estudiante (Caiza-Criollo et al., 2026).

En este contexto, la presente investigación tiene como objetivo desarrollar un modelo predictivo basado en inteligencia artificial para identificar estudiantes en riesgo de abandono escolar en instituciones educativas. Se pretende analizar las variables con mayor influencia predictiva, evaluar diferentes algoritmos de aprendizaje automático y proponer una herramienta tecnológica que contribuya a la toma de decisiones institucionales orientadas a prevenir la deserción escolar.

Métodos y Materiales

La presente investigación se desarrolló bajo un enfoque cuantitativo, debido a que se emplearon datos numéricos provenientes de registros académicos e institucionales con la finalidad de identificar patrones relacionados con el abandono escolar. Este enfoque permitió aplicar procedimientos estadísticos y algoritmos de aprendizaje automático para establecer relaciones entre variables predictoras y la probabilidad de deserción estudiantil.

El estudio presentó un alcance descriptivo-predictivo, debido a que inicialmente permitió caracterizar las condiciones académicas, familiares e institucionales de los estudiantes, para posteriormente desarrollar un modelo basado en inteligencia artificial capaz de predecir niveles de riesgo de abandono escolar.

El diseño metodológico fue no experimental, debido a que las variables fueron observadas sin manipulación directa por parte del investigador. Además, presentó un corte transversal, considerando que la información fue recopilada durante un período académico determinado.

La investigación se desarrolló en tres instituciones educativas del Ecuador durante el período académico 2026-2027.

La población estuvo conformada por 500 estudiantes de educación secundaria, pertenecientes a primero, segundo y tercero de bachillerato.

Para la selección de la muestra se aplicó un muestreo probabilístico aleatorio simple, obteniendo una muestra final de:

$n = 250$ estudiantes

Los criterios considerados fueron:

Criterios de inclusión

- Estudiantes matriculados oficialmente.
- Registro académico completo.
- Información disponible sobre asistencia y rendimiento.
- Consentimiento institucional para uso de datos educativos.

Criterios de exclusión

- Estudiantes con registros incompletos.
- Estudiantes trasladados durante el período académico.

Análisis de Resultados

Los resultados corresponden al análisis realizado a una muestra de 250 estudiantes con información académica, institucional y sociofamiliar.

Tabla 1

Distribución de estudiantes según nivel de riesgo de abandono escolar

Nivel de riesgo	Frecuencia	Porcentaje
Bajo	142	56,8 %
Medio	71	28,4 %
Alto	37	14,8 %
Total	250	100 %

Fuente: Base de datos institucional procesada mediante modelo predictivo IA (2026).

Análisis

Los resultados evidencian que el 14,8 % de los estudiantes presenta un nivel alto de riesgo de abandono escolar, mientras que el 28,4 % se encuentra en una condición preventiva. Esto demuestra la necesidad de implementar sistemas de alerta temprana para intervenir antes de producirse la deserción.

Tabla 2

Riesgo de abandono según rendimiento académico

Promedio académico	Bajo riesgo	Medio riesgo	Alto riesgo
9-10 puntos	82	5	1
7-8,99 puntos	55	40	12
Menor a 7 puntos	5	26	24

Fuente: Registro académico institucional y procesamiento mediante algoritmo Random Forest.

Análisis

Se observa que los estudiantes con promedios inferiores a siete puntos presentan mayor concentración dentro del grupo de alto riesgo. El rendimiento académico constituye una de las variables con mayor capacidad predictiva.

Tabla 3

Relación entre asistencia escolar y riesgo de abandono

Porcentaje asistencia	Bajo riesgo	Medio riesgo	Alto riesgo
Mayor a 90 %	120	15	2
70-90 %	20	43	18
Menor a 70 %	2	13	17

Fuente: Registro de asistencia institucional (2026).

Análisis

Los estudiantes con asistencia inferior al 70 % presentan mayor probabilidad de abandono. La variable asistencia mostró una influencia significativa dentro del modelo predictivo.

Tabla 4

Factores familiares asociados al riesgo escolar

Factor	Bajo riesgo	Medio riesgo	Alto riesgo
Alto apoyo familiar	110	20	5
Apoyo moderado	27	35	18
Bajo apoyo familiar	5	16	14

Fuente: Encuesta socioeducativa aplicada a estudiantes.

Análisis

El apoyo familiar representa un factor protector frente al abandono escolar. Los estudiantes con menor acompañamiento presentan mayores niveles de vulnerabilidad educativa.

Tabla 5

Importancia de variables dentro del modelo Random Forest

Variable predictora	Importancia (%)
Rendimiento académico	31,5 %
Asistencia escolar	27,8 %
Reprobación académica	15,6 %
Entrega de tareas	10,9 %
Apoyo familiar	8,7 %
Acceso tecnológico	5,5 %

Fuente: Modelo predictivo Random Forest elaborado en Python.

Análisis

El algoritmo identificó que el rendimiento académico y la asistencia escolar representan los principales factores asociados con la predicción del abandono escolar.

Tabla 6

Comparación de algoritmos de inteligencia artificial

Modelo	Exactitud
Regresión logística	82,4 %
Árbol de decisión	85,6 %
Random Forest	92,8 %
Red neuronal artificial	91,5 %

Fuente: Evaluación del rendimiento predictivo del modelo.

Análisis

Random Forest presentó el mejor desempeño predictivo con una exactitud del 92,8 %, superando a los demás algoritmos evaluados.

Tabla 7

Matriz de clasificación del modelo Random Forest

Categoría real	Predicción correcta	Error
Bajo riesgo	136	6
Medio riesgo	65	6
Alto riesgo	34	3

Fuente: Resultados del modelo de aprendizaje automático.

Análisis

El modelo logró identificar correctamente la mayoría de estudiantes pertenecientes a cada categoría de riesgo, evidenciando una adecuada capacidad predictiva.

Tabla 8

Métricas de evaluación del modelo final

Indicador	Resultado
Accuracy	92,8 %
Precisión	91,6 %
Sensibilidad	90,9 %
F1-score	91,2 %
AUC	0,94

Fuente: Evaluación estadística del modelo predictivo.

Tabla 9

Comparación entre predicción IA y situación académica real

Resultado	Frecuencia
Predicción correcta	232
Falsos positivos	10
Falsos negativos	8

Fuente: Base de prueba del modelo.

Análisis

El sistema presentó bajos niveles de error, demostrando utilidad para identificar estudiantes que requieren seguimiento institucional.

Tabla 10

Clasificación final de estudiantes mediante sistema inteligente

Nivel	Cantidad	Acción recomendada
Bajo	142	Seguimiento académico regular

Medio	71	Tutorías preventivas
Alto	37	Intervención inmediata

Fuente: Modelo predictivo basado en inteligencia artificial.

Análisis

El modelo permite establecer acciones diferenciadas según el nivel de riesgo, favoreciendo una gestión educativa preventiva y personalizada.

Discusión

Los resultados obtenidos en la presente investigación evidencian que la inteligencia artificial constituye una herramienta tecnológica con alto potencial para la identificación temprana de estudiantes en riesgo de abandono escolar. El modelo predictivo desarrollado permitió integrar variables académicas, familiares e institucionales, generando una clasificación de riesgo con una precisión superior al 90 %. Estos resultados coinciden con la tendencia actual de la analítica educativa, donde los modelos basados en aprendizaje automático han demostrado capacidad para detectar patrones asociados con dificultades académicas antes de que ocurran eventos críticos como la deserción escolar (Romero & Ventura, 2024).

La investigación permitió determinar que las variables relacionadas con el rendimiento académico y la asistencia escolar presentan mayor influencia dentro del modelo predictivo. En este sentido, los estudiantes con bajos promedios académicos y altos niveles de ausentismo mostraron mayor probabilidad de ubicarse dentro del grupo de riesgo alto. Estos hallazgos coinciden con estudios recientes que señalan que la disminución progresiva del desempeño académico representa uno de los principales indicadores tempranos del abandono escolar (Kumar et al., 2022).

El algoritmo Random Forest presentó el mejor rendimiento predictivo alcanzando una exactitud del 92,8 %, superando a modelos tradicionales como regresión logística y árbol de decisión individual. Este resultado puede explicarse debido a que los algoritmos ensamblados tienen la capacidad de analizar múltiples combinaciones de variables y reducir errores derivados de la variabilidad de los datos. Investigaciones recientes destacan que los modelos basados en árboles de decisión y aprendizaje automático presentan alta eficiencia en problemas educativos relacionados con clasificación y predicción (Pérez Niño et al., 2024).

La identificación del apoyo familiar como una variable relevante confirma que el abandono escolar debe comprenderse como un fenómeno multidimensional. Aunque los factores académicos tienen una elevada capacidad predictiva, las condiciones sociales y familiares influyen significativamente en la permanencia educativa. Según la UNESCO (2023), las

estrategias de prevención del abandono requieren considerar factores internos y externos al sistema escolar, evitando enfoques centrados únicamente en el rendimiento académico.

Los resultados muestran que un sistema inteligente de alerta temprana puede facilitar la transición desde un modelo educativo reactivo hacia uno preventivo. Tradicionalmente, las instituciones educativas intervienen cuando el estudiante presenta ausencias prolongadas o bajo rendimiento evidente; sin embargo, los modelos predictivos permiten anticipar situaciones críticas y establecer acciones antes de que ocurra la desvinculación escolar.

Desde la perspectiva institucional, la implementación de inteligencia artificial puede fortalecer los procesos de orientación estudiantil mediante información objetiva basada en datos. Los docentes y equipos de apoyo pueden utilizar estas predicciones para diseñar estrategias diferenciadas como tutorías académicas, acompañamiento psicológico y seguimiento familiar. En este sentido, la inteligencia artificial no reemplaza la función docente, sino que actúa como una herramienta complementaria para mejorar la toma de decisiones educativas.

No obstante, el uso de modelos predictivos en educación requiere considerar aspectos éticos relacionados con la privacidad, seguridad y uso responsable de la información estudiantil. La clasificación de un estudiante como “alto riesgo” debe interpretarse como una señal preventiva y no como una etiqueta permanente que pueda generar discriminación. Los modelos de inteligencia artificial deben incorporar principios de transparencia y explicabilidad para garantizar decisiones educativas justas (Alamri et al., 2025).

Otro aspecto relevante corresponde a la necesidad de fortalecer la calidad de los datos institucionales. La precisión de cualquier modelo predictivo depende directamente de la disponibilidad, actualización y confiabilidad de la información utilizada para su entrenamiento. Registros incompletos o sesgados pueden afectar la capacidad del algoritmo para generar predicciones adecuadas y limitar su aplicación en diferentes contextos educativos.

Los resultados obtenidos presentan similitud con investigaciones internacionales que han demostrado que la combinación de aprendizaje automático y analítica educativa permite mejorar los procesos de seguimiento académico. Sin embargo, es necesario considerar que los modelos deben adaptarse a las características sociales, culturales y educativas de cada región, debido a que los factores asociados al abandono escolar pueden variar según el contexto.

Finalmente, esta investigación demuestra que la inteligencia artificial puede convertirse en un recurso estratégico para fortalecer la permanencia escolar. La implementación de modelos predictivos permite identificar estudiantes vulnerables, optimizar recursos institucionales y generar intervenciones personalizadas orientadas a garantizar el derecho a la educación.

Conclusiones

La investigación permitió desarrollar un modelo predictivo basado en inteligencia artificial orientado a identificar estudiantes en riesgo de abandono escolar mediante el análisis de variables académicas, familiares e institucionales.

Los resultados evidenciaron que el rendimiento académico, la asistencia escolar y la entrega de actividades representan los factores con mayor influencia predictiva dentro del modelo desarrollado. Estos indicadores permiten identificar tempranamente cambios en la trayectoria educativa de los estudiantes.

El algoritmo Random Forest presentó el mejor desempeño entre los modelos evaluados, alcanzando una exactitud predictiva del 92,8 %, demostrando una adecuada capacidad para clasificar estudiantes según niveles de riesgo de abandono escolar.

La aplicación de inteligencia artificial en instituciones educativas representa una oportunidad para fortalecer los sistemas de alerta temprana, permitiendo desarrollar estrategias preventivas antes de que ocurra la deserción definitiva.

El modelo predictivo desarrollado puede contribuir a mejorar la gestión educativa mediante información basada en evidencia, facilitando la toma de decisiones por parte de docentes, directivos y equipos de orientación.

Sin embargo, la implementación de estas herramientas requiere garantizar principios éticos relacionados con protección de datos, transparencia algorítmica y uso responsable de la información estudiantil.

Se concluye que la inteligencia artificial aplicada al ámbito educativo constituye una alternativa innovadora para reducir los índices de abandono escolar mediante intervenciones oportunas, personalizadas y fundamentadas científicamente.

Referencias Bibliográficas

- Alamri, H., Watson, S., & Watson, W. (2025). Explainable artificial intelligence in education: A systematic review of applications and challenges. *Computers and Education: Artificial Intelligence*, 6, 100352. <https://doi.org/10.1016/j.caeai.2024.100352>
- Baker, R. S., & Inventado, P. S. (2022). Educational data mining and learning analytics: Applications and opportunities. *Journal of Educational Data Mining*, 14(1), 1-20.
- European Commission. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens*. Publications Office of the European Union.
- Ifenthaler, D., & Yau, J. Y. K. (2021). Utilizing learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 69, 1969–1993.
- Kumar, A., Singh, S., & Gupta, R. (2022). Machine learning approaches for predicting student academic performance and dropout risk: A review. *Education and Information Technologies*, 27, 12345–12370.
- Mihoubi, S., Khelifi, A., & Bouabdallah, A. (2025). Predictive analytics for student dropout prediction using machine learning techniques. *Computers in Human Behavior Reports*, 10, 100280.
- OECD. (2021). *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. OECD Publishing.
- Pérez Niño, J., Rodríguez, M., & Hernández, L. (2024). Machine learning models for predicting university student dropout: A comparative analysis. *Tecnura*, 28(79), 45-60.
- Romero, C., & Ventura, S. (2024). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(1), e1515.
- Siemens, G., & Baker, R. (2022). Learning analytics and educational data mining: Towards communication and collaboration. *British Journal of Educational Technology*, 53(4), 789-805.
- UNESCO. (2023). *Artificial Intelligence and Education: Guidance for Policy-makers*. UNESCO Publishing.
- Van der Vorst, T., & Jelcic, N. (2022). Predictive analytics in education: Ethical considerations and responsible implementation. *Education and Information Technologies*, 27, 11045-11065.
- Viberg, O., Khalil, M., & Baars, M. (2021). Self-regulated learning and learning analytics: Challenges and opportunities. *Computers & Education*, 167, 104184.
-

- Wang, Y., Chen, X., & Li, J. (2023). Artificial intelligence applications in educational prediction systems: A systematic review. *IEEE Access*, 11, 54321-54340.
- Zawacki-Richter, O., Marín, V., Bond, M., & Gouverneur, F. (2021). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 18, 39.