

DOI: 10.64325/IJIP.v3i1.56

Modelos de aprendizaje automático para la predicción de la deserción escolar en contextos rurales

Machine learning models for predicting school dropout rates in rural contexts

Oscar Iván Montiel Petro

University of Technology and education, Education, USA

<https://orcid.org/0009-0000-4287-0862>

oscar.montielp2024@uted.us

Corresponding author: oscar.montielp2024@uted.us

Citar como: Montiel Petro, O. I. (2026). Modelos de aprendizaje automático para la predicción de la deserción escolar en contextos rurales. International Journal of Innovation and Pedagogy, 3(1), 68-82. <https://doi.org/10.64325/IJIP.v3i1.56>

Recibido: 25-02-2026

Aceptado: 13-03-2026

Publicado: 15-05-2026

Editor: Mgtr. Daniel Roman-Acosta PhD. (c)

Copyright: © 2026 Oscar Iván Montiel Petro; Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0) 

ABSTRACT

This research systematizes the academic evidence on the use of machine learning models to predict school dropout rates in basic and secondary education in rural contexts (2015-2025). A systematic review was conducted following PRISMA 2020 guidelines, and Scopus, Web of Science, SciELO, and Redalyc were consulted. After applying eligibility criteria focused on rurality and vulnerable populations, 25 articles were selected. The results show the predominance of ensemble models, especially Random Forest and XGBoost, due to their performance with scarce and unbalanced tabular data. A shift in the most influential predictors is also observed: structural and geographic variables—home-school distance, maternal educational level, and agricultural cycles—outperform prior academic performance. In parallel, the explainability of models (XAI) emerges as a key requirement for implementing Early Warning Systems that guide timely decisions. In summary, machine learning outperforms traditional statistical approaches for detecting dropout risk in rural areas. However, its effectiveness depends on integrating contextualized data and translating algorithmic results into pedagogical actions that are understandable for teachers, administrators, and local authorities. There are also reported concerns about the quality and availability of administrative data, as well as about ethics, privacy, and bias in rural communities.

Keywords: school dropout, machine learning, rural education, predictive models, systematic review, artificial intelligence.

RESUMEN

Esta investigación sistematiza la evidencia académica sobre el uso de modelos de aprendizaje automático para predecir la deserción escolar en educación básica y media en contextos rurales (2015-2025). Se realizó una revisión sistemática siguiendo PRISMA 2020 y se consultaron Scopus, Web of Science, SciELO y Redalyc. Tras aplicar criterios de elegibilidad centrados en ruralidad y poblaciones vulnerables, se seleccionaron 25 artículos. Los resultados muestran el predominio de modelos de ensamble, especialmente Random Forest y XGBoost, por su desempeño con datos tabulares escasos y desbalanceados. También se observa un cambio en los predictores más influyentes: variables estructurales y geográficas — distancia hogar-escuela, nivel educativo materno y ciclos agrícolas— superan al rendimiento académico previo. En paralelo, la explicabilidad de modelos (XAI) emerge como requisito clave para implementar Sistemas de Alerta Temprana que orienten decisiones oportunas. En síntesis, el aprendizaje automático supera a enfoques estadísticos tradicionales para detectar riesgo de deserción en la ruralidad; sin embargo, su efectividad depende de integrar datos situados y de traducir resultados algorítmicos en acciones pedagógicas comprensibles para docentes, directivos y autoridades territoriales. Se reporta además una preocupación por la calidad y disponibilidad de datos administrativos, así como por la ética, la privacidad y el sesgo en comunidades rurales.

Palabras clave: deserción escolar, aprendizaje automático, educación rural, modelos predictivos, revisión sistemática, inteligencia artificial.

INTRODUCCIÓN

La deserción escolar en educación básica y media es una realidad, así lo indican numerosos estudios sobre el tema en diversas partes del mundo, por lo tanto, representa uno de los desafíos más preocupantes para los sistemas educativos a nivel mundial. Mientras las tasas globales de abandono han mostrado descensos moderados en las últimas décadas, las brechas entre zonas urbanas y rurales se mantienen prácticamente intactas (Montes y Luna, 2024; UNESCO, 2021).

En estos contextos, la deserción alcanza niveles que duplican o triplican los promedios nacionales, perpetuando ciclos de desigualdad y pobreza que afectan desproporcionadamente a las comunidades vulnerables (Rodríguez et al., 2023; CEPAL, 2020; Retamal et al., 2021) y amplían considerablemente la brecha social en estos espacios marginados y muchas veces olvidados.

Esta situación adquiere particular relevancia cuando se consideran las limitaciones estructurales y socioeconómicas propias de las zonas rurales, semi-rurales, indígenas y campesinas, donde la escasez de recursos, la distancia a los centros educativos y la necesidad de incorporarse tempranamente al trabajo agrícola actúan como barreras adicionales para la continuidad académica (Huatangari et al., 2020). La mayoría de los padres de familia de los jóvenes del sector rural, prefieren ver a sus hijos trabajando, colaborando con los gastos del hogar, antes que verlos estudiar, ya que consideran que el estudio es una pérdida de tiempo.

Los estudiantes que abandonan prematuramente el sistema educativo no solo interrumpen su desarrollo personal y académico, sino que también enfrentan mayores probabilidades de desempleo, precariedad laboral y exclusión social en el futuro (Andrade-Girón et al., 2023). Para Galván Mora (2020), en estas condiciones rurales, los efectos se amplifican por la escasez de oportunidades laborales, la migración forzada y los ciclos constantes de vulnerabilidad social.

Las causas de la deserción escolar en entornos rurales son totalmente diferentes de las documentadas en los espacios urbanos. Factores como la distancia física a los establecimientos educativos, la necesidad de contribuir a economías familiares basadas en la agricultura y la ganadería, la poca oferta educativa en niveles superiores de educación media y las barreras socioculturales relacionadas a sistemas educativos poco pertinentes para comunidades indígenas y campesinas configuran un escenario causal particular que exige enfoques analíticos específicos para su comprensión y abordaje (Guzmán et al., 2021).

La naturaleza multicausal del fenómeno de la deserción escolar hace que sea mucho más difícil su erradicación. No es posible darle un mismo tratamiento a un niño que deserta de su escuela porque no se siente a gusto en ella y no entiende ni le emociona las asignaturas que recibe que a un niño que deserta porque la economía en su hogar no permite gastos extras como transporte y útiles escolares.

A esto se suman condiciones estructurales como la precariedad de la infraestructura educativa, la falta de recursos tecnológicos, la baja especialidad docente en aulas multigrado y la violencia que por años han afectado a estas poblaciones, limitan el acceso a una educación de calidad en estas regiones (Caballero, 2023; Galabay-Cajas y Álvarez-Lozano, 2021). A pesar de los intentos gubernamentales de reducir la brecha urbano-rural, muchas veces no resulta suficiente.

No basta con dotar a las instituciones educativas con recursos digitales, internet y computadores si no que es necesario masificar un plan de alfabetización docente, debido a que en un alto porcentaje el docente rural vive en la población donde labora y no encuentra los medios y las facilidades para capacitarse y mantenerse actualizado.

Ante esta complejidad, los sistemas tradicionales de identificación de estudiantes en riesgo resultan ineficaces. Los enfoques basados en datos fijos de asistencia o rendimiento académico frecuentemente fallan en capturar la dinámica multidimensional de los factores de riesgo en entornos rurales, donde variables socioeconómicas, geográficas y culturales interactúan de manera compleja (Jiménez-Rogel y Campoverde-Moscol, 2024; Torres, 2025). Esta limitación ha impulsado la búsqueda de metodologías más predictivas que permitan anticipar situaciones de riesgo e integrar múltiples variables para predecir el abandono escolar con mayor precisión (Castillo-Peña, 2021; Rincón et al., 2022).

Algunas posibilidades para abordar el tema de la deserción escolar en el entorno rural consisten en analizar la problemática desde una perspectiva retrospectiva y con los datos que se tienen a la mano de los estudiantes desertores, tratar de determinar cuáles fueron las causas de la deserción y compararla con la realidad de los estudiantes actuales para sí identificar que estudiante se encuentra en riesgo real de deserción y tratar de implementar

estrategias para evitar que esto ocurra. Realizar esta tarea de una forma manual es realmente un proceso complejo y un tanto inexacto, por tal motivo es importante analizar las nuevas tecnologías útiles para este tipo de estudios.

Una alternativa viable, estructurada, que ha tomado fuerza en los últimos años, es el aprendizaje automático (Machine Learning ML). Aunque no ha sido muy utilizada para abordar la problemática de la deserción escolar en entornos rurales, la experiencia en educación superior y contextos urbanos permite visualizar que es una alternativa seria y prometedora, teniendo en cuenta su efectividad trabajando con grandes cantidades de datos, tal como lo indican (Castrillón-Gómez et al., 2020; Nagy y Molontay, 2023), “los modelos de machine learning (ML) permite identificar patrones complejos, procesar grandes volúmenes de datos heterogéneos y generar predicciones con mayor precisión que los métodos estadísticos tradicionales”.

Algoritmos como Random Forest, Support Vector Machines y redes neuronales han demostrado superioridad en la detección de patrones no lineales dentro de grandes conjuntos de datos, permitiendo incorporar variables socioeconómicas y geográficas específicas de los entornos rurales que los modelos tradicionales no logran procesar eficazmente.

Conociendo estas técnicas de análisis de datos, resulta posible pensar en que pueden ser muy útiles para analizar el problema de la deserción escolar en contextos rurales, sin embargo, la mayor parte de la investigación en ML aplicado a la deserción escolar se ha desarrollado en contextos urbanos y universitarios. Los estudios en educación básica y media son menos frecuentes, y aquellos específicamente centrados en poblaciones rurales, indígenas o campesinas constituyen una minoría dentro de la producción científica actual (Restrepo, 2022).

Esta concentración geográfica y poblacional genera interrogantes sobre la transferibilidad de los modelos desarrollados en entornos urbanos hacia realidades rurales con características estructurales y socioeconómicas marcadamente distintas (Caballero, 2023). ¿Son válidas y aplicables las variables predictoras identificadas en entornos urbanos para contextos rurales? ¿Qué adaptaciones metodológicas requieren los modelos ML cuando funcionan con datos escasos, fragmentados o culturalmente situados?

El desafío que significa abordar el problema de la deserción escolar en la ruralidad utilizando tecnologías como el machine learning adolece de investigaciones sistemáticas que nos den luces de la realidad de este fenómeno. La ausencia de síntesis sistemáticas que aborden específicamente esta brecha limita tanto el desarrollo teórico como la implementación práctica de sistemas predictivos en zonas rurales. Revisiones previas sobre ML y deserción escolar han incluido estudios rurales de manera esporádica o las han incluido dentro de categorías generales sin profundizar en las particularidades que definen estos contextos (Andrade-Girón et al., 2023; Gamboa-Cruzado et al., 2023).

Las investigaciones se han centrado en educación superior, donde los estudiantes tienen realidades diferentes y el contexto varía significativamente de la educación básica y media en el sector rural, razón por la cual no resulta óptimo basarse en estos estudios para abordar el caso en un contexto no urbano y no es posible encontrar factores determinantes y disponibilidad de datos para el modelo predictivo (Andrade-Girón et al., 2023).

La búsqueda de documentos académicos sobre ML aplicada a la deserción en contextos rurales ha sido extensa, numerosas bases de datos académicas llevan a concluir que hasta donde se ha podido constatar, no existen revisiones sistemáticas que aborden de manera exclusiva la aplicación de ML y consoliden la evidencia para la predicción de la deserción escolar en contextos rurales de educación básica y media (Forero-Corba y Bennásar, 2023; Salce, 2020).

Esta situación más que un problema, resulta ser una oportunidad para realizar un estudio completo, detallado y contextualizado de la realidad de la deserción escolar en las escuelas del área rural. Esta revisión busca llenar este vacío. Sistematizar la evidencia disponible permitirá identificar qué modelos han sido empleados, qué variables predictoras resultan relevantes en estos contextos, cuáles son las métricas de desempeño alcanzadas y qué limitaciones metodológicas persisten.

Los resultados pueden orientar el diseño de políticas educativas basadas en evidencia, informar el desarrollo tecnológico de plataformas de alerta temprana contextualizadas y trazar una agenda de investigación que priorice poblaciones históricamente subrepresentadas en la literatura sobre analítica educativa. El objetivo de esta revisión es sistematizar la literatura existente sobre modelos de aprendizaje automático aplicados a la predicción de

deserción escolar en estudiantes de educación básica y media en entornos rurales, semi-rurales, indígenas y campesinos, publicados entre 2015 y 2025.

Concretamente, se busca, caracterizar los modelos ML empleados y su desempeño predictivo, identificar las variables predictoras, describir los contextos rurales estudiados y sus metodologías y analizar limitaciones y vacíos en la evidencia actual. Por tanto, la pregunta que orienta esta revisión es: ¿Qué modelos de aprendizaje automático se han empleado para predecir la deserción escolar en estudiantes de básica y media en contextos rurales, y cuáles son sus variables predictoras, métricas de desempeño y limitaciones metodológicas? Responder esta pregunta contribuye no solo al avance en analítica educativa, sino también a la posibilidad real de reducir inequidades que afectan a las poblaciones más vulnerables.

METODOLOGIA

Esta investigación se desarrolló como una revisión sistemática de literatura siguiendo los pasos establecidos en la declaración PRISMA 2020 (Page et al., 2021). El protocolo de revisión fue registrado previamente en Open Science Framework para garantizar la transparencia y reproducibilidad del proceso de selección y análisis de los estudios (Gamboa-Cruzado et al., 2023). La revisión sistemática se eligió diseño apropiado porque permite sintetizar de manera rigurosa y replicable la evidencia disponible sobre una pregunta de investigación específica, minimizando sesgos mediante procesos explícitos y reproducibles (Higgins et al., 2019).

Los criterios de inclusión y exclusión se definieron siguiendo el marco PICOS adaptado para revisiones de estudios predictivos (Moons et al., 2014). El periodo de búsqueda se delimitó desde enero de 2015 hasta enero de 2025 para asegurar la inclusión de estudios recientes que reflejen los avances actuales en aprendizaje automático aplicado a entornos rurales (Romero y Ventura, 2020; Andrade-Girón et al., 2023; Tete et al., 2022). La población de interés correspondió a estudiantes de educación básica y media en contextos rurales, la intervención a la aplicación de modelos de aprendizaje automático, y los resultados a las variables predictoras identificadas, métricas de desempeño alcanzadas y limitaciones metodológicas reportadas.

Se aceptaron publicaciones en español, inglés o portugués, idiomas predominantes en la producción científica latinoamericana y global sobre educación rural. El tipo de estudio se restringió a investigaciones empíricas con diseños cuantitativos o mixtos que incluyeran desarrollo, validación o aplicación de modelos de aprendizaje automático, excluyendo revisiones de literatura, estudios puramente teóricos, editoriales, comentarios y opiniones.

En cuanto al contexto geográfico, se consideraron instituciones educativas ubicadas en zonas clasificadas como rurales, semi-rurales, rurales dispersas, comunidades indígenas o comunidades campesinas según definiciones oficiales de cada país o categorización explícita de los autores. La intervención tecnológica debía implicar la aplicación de al menos un algoritmo de aprendizaje automático supervisado o no supervisado para predecir deserción escolar, abandono escolar o riesgo de desvinculación. Se incluyeron técnicas como árboles de decisión, Random Forest, Support Vector Machines, regresión logística, redes neuronales, Gradient Boosting, k-Nearest Neighbors, Naive Bayes y análisis discriminante, entre otros.

La variable de resultado considerada fue deserción escolar, abandono escolar, retiro del sistema educativo o riesgo de desvinculación como variable dependiente principal del modelo predictivo. Se aceptaron como tipos de publicación artículos en revistas científicas indexadas, artículos de conferencias arbitradas con revisión por pares, capítulos de libro en editoriales académicas y tesis doctorales disponibles en repositorios institucionales.

Los criterios de exclusión comprendieron estudios centrados exclusivamente en educación superior, educación técnica post-secundaria o formación de adultos, así como investigaciones desarrolladas únicamente en contextos urbanos o metropolitanos sin componente rural. También se excluyeron estudios que utilizaran únicamente métodos estadísticos tradicionales sin incorporar algoritmos de aprendizaje automático, publicaciones sin acceso a texto completo tras solicitud a autores o bibliotecas institucionales, estudios duplicados o publicaciones múltiples del mismo conjunto de datos sin aportes metodológicos adicionales, e investigaciones que no especificaran el contexto geográfico de la muestra o que mezclaran datos urbanos y rurales sin desagregación.

Se consultaron bases de datos multidisciplinarias y especializadas en ciencias sociales y educación, incluyendo Scopus, Web of Science, ERIC, IEEE Xplore, Science Direct, Redalyc, Google Scholar, Scielo, Dialnet y DOAJ, para garantizar una cobertura amplia de la

literatura académica y técnica relevante (Haddaway et al., 2015; Gamboa-Cruzado et al., 2023; Mariñas et al., 2022).

Las ecuaciones de búsqueda se construyeron combinando tres bloques temáticos mediante operadores booleanos (AND, OR): “deserción escolar”, “school dropout”, “student attrition”, “early school leaving”, “aprendizaje automático”, “machine learning”, “data mining”, “predictive model”, “early warning system”, “aprendizaje automático”, “modelo predictivo”, y “contextos rurales”, “rural education”, “rural school”, “educación rural”, “educación al campesino”, “educación indígena”.

Tipo y diseño de investigación: La investigación corresponde a una revisión sistemática de la literatura; uso de bases de datos (Scopus, Web of Science, ERIC, SciELO, Redalyc); matriz de extracción de datos en Excel.

Población: Estudios científicos relacionados con el tema de machine learning en la deserción escolar en el contexto rural.

Muestra: 25 artículos seleccionados mediante protocolo PRISMA

Normas/estándares aplicados: Protocolo PRISMA 2020; uso de descriptores controlados (Tesauros UNESCO y ERIC)

Técnicas de recolección de datos: Búsqueda bibliográfica sistemática; selección y cribado (PRISMA); matriz de extracción de datos

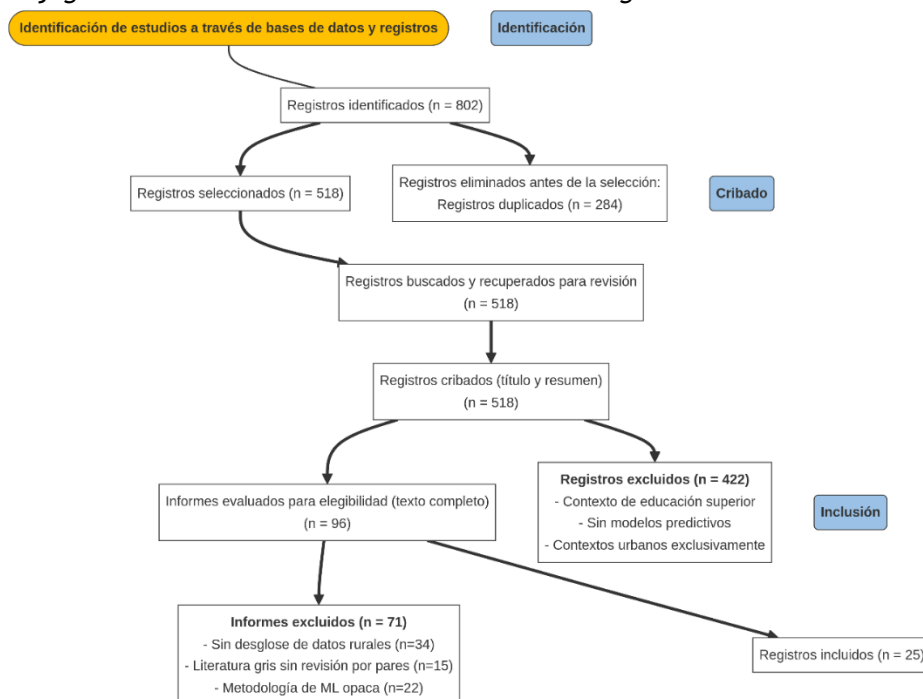
Técnicas de análisis de datos: Análisis temático (Braun y Clarke (2006)); codificación inductiva y deductiva; síntesis cualitativa; análisis descriptivo.

RESULTADOS

La Figura 1 presenta el diagrama de flujo que detalla el número de registros identificados, incluidos y excluidos, así como las razones específicas de exclusión en cada etapa del proceso.

Figura 1

Flujograma del Proceso de Selección de Estudios Según la Declaración PRISMA 2020)



Nota: El diagrama expone las fases de identificación, cribado, elegibilidad e inclusión. De un total inicial de 802 registros, se seleccionaron 25 estudios para la síntesis cualitativa. Adaptado de "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", por M. Page et al., 2021, BMJ.

El proceso de selección siguió el flujo de trabajo establecido por la declaración PRISMA 2020, el cual estipula las fases de identificación, cribado, elegibilidad e inclusión. Durante la primera fase se recuperaron un total de 802 registros. En la fase de cribado tras usar un gestor bibliográfico se eliminaron 284 duplicados, quedando 518 registros para su evaluación.

Posteriormente, se procedió a la lectura de títulos y resúmenes para excluir aquellos estudios que no cumplieran con los criterios de elegibilidad establecidos, lo que resultó en la exclusión de 422 registros y la selección de 96 estudios. Estos 96 registros fueron sometidos a una lectura de texto completo, resultando en la inclusión definitiva de 25 artículos para el análisis cualitativo y de síntesis. La extracción de la información se realizó mediante una matriz estandarizada en Microsoft Excel que permitió registrar sistemáticamente datos como autor, año, país, nivel educativo, tipo de contexto rural, algoritmos empleados, variables predictoras, métricas de desempeño y limitaciones reportadas (Wolff et al., 2019).

Tras aplicar los criterios de elegibilidad y completar el proceso de selección según la guía PRISMA 2020, se identificaron 25 estudios primarios que constituyen la muestra final de esta revisión sistemática. Basado en las fuentes, surgieron cuatro categorías temáticas que agrupan los hallazgos: Arquitectura Técnica y Algoritmos de ML, Clasificación de Variables Predictoras (Determinantes), Especificidades del Contexto Rural y Explicabilidad (XAI), e Impacto en Políticas.

Arquitectura Técnica y Algoritmos de ML

Los estudios analizados en esta investigación bibliográfica muestran una tendencia en el uso de algoritmos de aprendizaje supervisado como Random Forest y XGBoost. Buscando causas de esta predominancia sobre otro tipo de algoritmos como los no supervisados, se logró determinar que estos resultan más adecuados para este tipo de estudios, debido a su rendimiento superior en términos de precisión y sensibilidad para la clasificación de estudiantes en riesgo de deserción en zonas rurales (Andrade-Girón et al., 2023; Parra-Sánchez et al., 2023).

Estos modelos de conjunto (ensemble learning), superan consistentemente a clasificadores individuales como la regresión logística o los árboles de decisión simples al manejar conjuntos de datos desbalanceados típicos de entornos educativos rurales (Marcolino et al., 2025; Hooshyar et al., 2025). Otras técnicas frecuentemente reportadas incluyen Support Vector Machines, redes neuronales artificiales y k-Nearest Neighbors, las cuales se aplican tanto en modelos individuales como en combinaciones híbridas para optimizar la capacidad predictiva (Andrade-Girón et al., 2023; Gamboa-Cruzado et al., 2023).

La literatura indica que la selección de algoritmos se fundamenta en su capacidad para manejar datos faltantes y relaciones no lineales entre variables socioeconómicas y académicas propias de entornos rurales (Gamboa-Cruzado et al., 2023; Guzmán et al., 2021). Entre las técnicas menos frecuentes pero documentadas se encuentran algoritmos avanzados como Causal Forest, Convolutional Neural Networks, Logistic Model Trees y Gaussian Process Regression, los cuales se emplean en contextos específicos para abordar complejidades adicionales en los datos (Forero-Corba y Bennásar, 2023).

Fuentes coreanas destacan que el uso de técnicas como SMOTE (sobremuestreo de minorías) ha permitido mitigar el problema del desbalance de clases en los conjuntos de datos, mejorando significativamente la capacidad de generalización de los modelos predictivos en poblaciones estudiantiles rurales donde los casos de deserción son numéricamente inferiores a los de permanencia (Álamo, 2024; Niño et al., 2024). Esos estudios llevados a cabo en otras zonas geográficas son una gran contribución para el desarrollo de sistemas adaptados a los contextos y realidades con los que se está en interacción.

En cuanto a las métricas más allá de la precisión, los estudios más recientes priorizan el AUC-ROC (Area Under the Curve) y la sensibilidad, ya que proporcionan una evaluación más robusta del rendimiento al detectar correctamente a los pocos alumnos en riesgo de deserción, minimizando así los falsos negativos que podrían pasar por alto a estudiantes vulnerables (Marcolino et al., 2025; Melo y Souza, 2023). Resulta importante identificar los posibles falsos negativos que pueden hacer caer en error al sistema de alerta para ofrecer una respuesta más profunda, certera y de esta forma demostrar esta diferencia con respecto a los sistemas tradicionales basados en estadística y probabilidades.

Clasificación de Variables Predictoras

Las variables predictoras encontradas en la literatura se agrupan en dimensiones individuales, académicas, socioeconómicas e institucionales/contextuales que influyen en la decisión de abandonar el sistema educativo en las escuelas rurales.

En la dimensión individual y académica resalta el rendimiento académico, la asistencia a clases y el historial de reprobación como los indicadores más fuertes para la detección

temprana de riesgo de abandono escolar (Tong y Li, 2024; Villarrasa-Sapiña et al., 2024). La extra-edad y pertenecer a grupos étnicos minoritarios se identifican adicionalmente como factores de riesgo, asociados a menudo con barreras lingüísticas y culturales que dificultan la adaptación de estas poblaciones al entorno escolar (Andrade-Girón et al., 2023).

El hecho de no sentirse bien recibido e identificado con sus compañeros de escuela lleva al estudiante a sentirse aislado y poco valorado, lo cual lo lleva en muchas ocasiones a pensar en el retiro de la escuela como una posibilidad de escape a aquellas situaciones que le incomodan y atormentan.

En la dimensión socioeconómica y familiar, la situación económica de los padres, el nivel educativo de estos y la estructura familiar surgen como determinantes críticos, dado que la necesidad de muchos estudiantes al trabajo y la falta de recursos materiales impactan directamente en la continuidad educativa (Zomaya, 2016). En este sentido, Fitriana et al. (2024), señala que la baja escolaridad de la madre puede aumentar de manera considerable el riesgo de deserción escolar en comparación con otros niveles educativos maternos.

En el ámbito institucional y contextual, elementos como la distancia física al centro educativo, el clima escolar, la disponibilidad de transporte y la calidad de la infraestructura escolar se reconocen como barreras estructurales que restringen el acceso y la permanencia estudiantil en zonas rurales (Mínguez-Martínez et al., 2024; Ocen et al., 2025). De igual forma, la baja conectividad, los ciclos de cosecha y la exposición a violencia o conflicto armado también se identifican como factores externos que agravan la vulnerabilidad de los estudiantes y obligan a la interrupción de sus estudios (Rincón et al., 2022). Todos estos factores anteriormente mencionados muestran que la multi causalidad de la deserción escolar hace necesario una revisión detallada y a conciencia para saber dónde nos encontramos y cuál es la ruta recomendada para abordar esta problemática que directa o indirectamente afecta a la mayoría de las personas en los territorios.

Por último, el cambio de institución al pasar de primaria a secundaria y la disponibilidad de grados superiores es un punto crítico que incrementa la deserción escolar debido a la falta de continuidad en la oferta educativa dentro de las mismas comunidades rurales (Guzmán et al., 2021; Roman, 2016). Es fundamental que las poblaciones alejadas de los centros urbanos cuenten con la continuidad desde el preescolar hasta el grado XI, este factor reduciría considerablemente la deserción en estudiantes que van avanzando a grados superiores.

Especificidades del Contexto Rural

La literatura muestra que la ruralidad no es solo una variable, sino un entorno que condiciona de manera sistemática la educación de la población rural aumentando el estado de vulnerabilidad. A diferencia del entorno urbano, en el campo la deserción está ligada a barreras geográficas y de acceso, a los ciclos productivos y a las condiciones de seguridad derivadas de conflictos armados en los territorios.

La dispersión poblacional y las largas distancias hasta los centros educativos constituyen barreras estructurales que dificultan el acceso diario y la permanencia de los estudiantes en el sistema escolar (Bayona-Rodríguez et al., 2022; Orozco y Rubio, 2019). El traslado la mayoría de las veces se vuelve oneroso y físicamente agotador para los estudiantes, lo que incrementa la probabilidad de abandono, especialmente cuando se combina con la necesidad de incorporarse al mercado laboral para apoyar la economía familiar (Rincón et al., 2022; Roman, 2016).

La necesidad de formación y lograr títulos profesionales no parece ser algo que inquiete a muchos estudiantes rurales. Adicionalmente, la marcada brecha digital de acceso a internet y carencia de servicios básicos como electricidad y agua potable limita las posibilidades de aprendizaje y el desarrollo de competencias digitales necesarias en la educación actual (Leal, 2019).

Reducir la denominada brecha digital, es un factor clave para los estudiantes rurales, puesto que ellos demuestran más atención y cercanía con herramientas digitales actuales, las cuales muchas veces llegan a las instituciones educativas, pero por diversos factores, entre ellos, la capacitación al docente, no se les da el uso adecuado y necesario para lograr alcanzar un aprendizaje significativo

Por otro lado, la pobreza estructural típica de las áreas rurales, el trabajo infantil y la movilidad estudiantil ligada a los ciclos de cosecha, así como la exposición a riesgos como el tráfico de drogas y el conflicto armado, se identifican como factores determinantes que

obligan a los estudiantes a abandonar sus estudios para contribuir a la subsistencia del hogar (Cilleros et al., 2017; S. et al., 2021).

Los estudiantes muchas veces abandonan sus estudios para contribuir al sustento familiar o para generar ingresos propios, situación que se agrava cuando los padres poseen bajos niveles de escolaridad, lo cual reduce las expectativas educativas familiares y la valoración del estudio como mecanismo de ascenso social (Rincón et al., 2022). Resulta para muchos padres de familia de entornos rurales mucho más conveniente que sus hijos trabajen en oficios varios en su lugar de origen y no arriesgarse a buscar nuevas y mejores oportunidades en otros horizontes.

Finalmente, el predominio de escuelas multigrado en las zonas rurales, donde un único docente debe atender simultáneamente a estudiantes de diferentes niveles educativos, representa un desafío pedagógico y logístico que impacta negativamente en la calidad del aprendizaje y en la continuidad educativa (Caballero, 2023). Esta configuración escolar, sumada a la escasez de recursos didácticos y tecnológicos, así como a la limitada preparación docente para gestionar aulas heterogéneas, genera desventajas significativas en relación con el contexto urbano que se traducen en bajos niveles de rendimiento académico y aumentan la probabilidad de deserción, principalmente en la educación media (Franco et al., 2020; Zambrano-Trujillo et al., 2022).

A continuación, se presenta una tabla 1, la cual sintetiza los autores más relevantes citados en los hallazgos de la revisión sistemática, destacando sus principales aportes en cuanto a modelos de aprendizaje automático y variables predictoras en contextos rurales.

Tabla 1

Principales autores y hallazgos sobre modelos de aprendizaje automático y variables predictoras en la educación rural (2015-2025).

Autor(es)	Año	Hallazgos / Contribuciones Relevantes
Quiñones Huatangari Carrasco Vega	2020 y	Argumentan que en zonas rurales la deserción es un fenómeno primordialmente social más que pedagógico, requiriendo modelos ajustados a esa realidad.
Galabay-Cajas Álvarez-Lozano	y 2021	Enfatizan la necesidad de incorporar "datos situados" (variables culturales, lenguas nativas y acceso a servicios) en los modelos para comunidades indígenas.
Bayona-Rodríguez et al.	2022	Documentan cómo la dispersión poblacional y las largas distancias actúan como barreras estructurales que incrementan el abandono escolar.
Andrade-Girón et al.	2023	Demuestran la superioridad de los modelos de ensamble (Random Forest y XGBoost) sobre modelos simples. Identifican la sobriedad y la pertenencia a grupos étnicos como factores de riesgo clave.
Fitriana et al.	2024	Identifican el bajo nivel educativo de la madre como un predictor crítico que aumenta significativamente el riesgo de deserción en comparación con otros factores.
Tong y Li; Villarrasa-Sapiña et al.	2024	Confirman el rendimiento académico y la asistencia como los indicadores individuales más fuertes para la detección temprana.
Bulut et al.	2024	Sostienen que la sensibilidad debe primar sobre la precisión general, dado que no detectar a un estudiante en riesgo impide la intervención oportuna.
Marcolino et al.	2025	Resaltan la eficacia de los modelos de conjunto para manejar datos desbalanceados y proponen priorizar las métricas AUC-ROC y sensibilidad para detectar alumnos vulnerables.
Khatun et al.	2025	Destacan la robustez de los métodos de ensamble para manejar valores faltantes en datos rurales y advierten sobre el alto costo social de los falsos negativos.
Rincón et al.	2022, 2023	Analizan factores externos como los ciclos de cosecha, la conectividad y la violencia; señalan que un modelo que no detecta al desertor pierde su valor social.

DISCUSIÓN

Los resultados de esta revisión muestran evidencia nueva, aunque limitada, sobre el potencial de los modelos de aprendizaje automático para predecir y prevenir la deserción escolar en entornos rurales. Los hallazgos nos indican que la efectividad de estos modelos no está garantizada únicamente por la precisión de los algoritmos, sino en la capacidad de integrar variables propias del entorno rural que la literatura urbana suele pasar por alto.

En este aspecto, los resultados muestran una tendencia hacia algoritmos de ensamble como Random Forest y XGBoost. Estos métodos han manifestado obtener precisiones superiores a los modelos tradicionales, coincidiendo con lo reportado por Andrade-Girón et al. (2023), quienes apuntan que, para datos heterogéneos, estas técnicas superan en estabilidad a las redes neuronales. En los espacios rurales, donde la limpieza de datos es una lucha continua debido a la escasez y fragmentación de información, la robustez de los métodos de ensamble resulta decisivo para manejar los valores faltantes sin sacrificar el rendimiento predictivo (Khatun et al., 2025; Psyridou et al., 2024).

En muchas ocasiones la información que tienen las instituciones educativas no se encuentra de una forma totalmente organizada y sistematizada, por lo tanto, es importante que el sistema algorítmico utilizado sea el idóneo para estos contextos y realidades educativas del entorno rural.

Sin embargo, se debe destacar en que la precisión general no es la métrica más sustancial. En situaciones de deserción, un falso negativo (un estudiante en riesgo no detectado) es mucho más costoso que un falso positivo, ya que no identificar a un estudiante en riesgo implica perder la oportunidad de intervenir a tiempo, mientras que una falsa alarma solo conlleva la asignación de recursos preventivos innecesarios (Bulut et al., 2024; Khatun et al., 2025). Por tanto, los futuros modelos deben priorizar la sensibilidad y el F1-score para ser realmente útiles en la prevención. (Rincón et al., 2023), señala que un modelo que no prioriza a él desertor pierde su valor social, perpetuando la invisibilidad de los estudiantes más vulnerables.

Un hallazgo crítico en esta revisión es que la mayoría de los modelos tienden a ser generalistas y no acogen las situaciones propias del campo. Mientras que en contextos urbanos el rendimiento académico es el principal predictor, en la ruralidad factores como la distancia entre el hogar y la escuela, el trabajo infantil agrícola y la educación de la madre juegan un papel mucho más importante. La ruralidad no puede ser vista como un riesgo en sí, sino como un contexto que requiere modelos ajustados a la realidad. Estos resultados apoyan la tesis de (Quiñones Huatangari y Carrasco Vega, 2020), quienes argumentan que en zonas rurales la deserción es un fenómeno más social que pedagógico.

La inclusión de estudios centrados en comunidades indígenas (Awajún y Wampis) y zonas post-conflicto es de destacar en la literatura sobre analítica educativa, pues, los modelos tradicionales suelen fallar ante estas comunidades. Por lo tanto, es imperativo que la investigación futura en aprendizaje automático para la ruralidad incorpore lo que (Galabay-Cajas y Álvarez-Lozano, 2021), denominan datos situados, es decir, información que refleje las variables culturales y socioeconómicas específicas, como las lenguas nativas o el acceso a servicios públicos de estas comunidades indígena y campesinas.

La efectividad de un sistema de alerta temprana en estas condiciones depende de su integración con sistemas de gestión nominales como lo es el SIMAT en Colombia. A pesar de los avances técnicos en las últimas décadas y del desarrollo de la inteligencia artificial, aún existe un desbalance entre velocidad de generación de datos y la capacidad real del estado para actuar sobre ellos en el territorio rural.

Finalmente, es necesario discutir sobre el auge de técnicas como SHAP y LIME mencionadas en las fuentes consultadas más recientes. Estas técnicas representan el avance más prometedor para la implementación de estos modelos en el campo. Como se observó en los hallazgos, el uso de cajas negras ha limitado la adopción tecnológica por parte de los docentes y directivos rurales.

La posibilidad de que un profesor entienda, mediante un lenguaje claro, por qué un estudiante ha sido marcado como en riesgo, humaniza el algoritmo y facilita la prevención de deserción. Por ende, un modelo que solo entrega una probabilidad de abandono es insuficiente para un director rural; el sistema debe explicar por qué el estudiante está en riesgo para que la intervención sea oportuna.

En consecuencia, los resultados de esta revisión refuerzan la idea de que el aprendizaje automático no debe reemplazar el juicio del docente, sino fortalecerlo. El impacto en las

políticas educativas solo será real si los modelos predictivos se integran en plataformas de gestión escolar que sean accesibles incluso en condiciones de baja conectividad.

CONCLUSIONES

Esta revisión sistemática confirma, el potencial que tienen los modelos de aprendizaje automático para prevenir la deserción escolar en contextos rurales. Las investigaciones han pasado de análisis descriptivos y sociológicos a orientaciones predictivas basadas en inteligencia artificial y Machine Learning, pasando de juicios retrospectivos a la prevención anticipada, fortaleciendo a la ciencia de los datos como una herramienta fundamental para la gestión en la educación.

Los resultados ratifican en primer lugar que los algoritmos de ensamble, en especial Random Forest y XGBoost, instituyen las arquitecturas más robustas y eficientes para el tratamiento de datos en la educación rural. Sin embargo, la métrica de sensibilidad de estos modelos debe primar sobre la precisión general, ya que el costo social de no detectar a un estudiante en riesgo de abandonar la escuela supera de manera considerable las implicaciones de una falsa alarma.

En segundo lugar, la evidencia indica un cambio de jerarquía de los predictores de abandono escolar. En los contextos rurales y comunidades en especial indígenas analizadas, las variables socioeconómicas y culturales, la distancia al centro educativo y el trabajo infantil, tienen un peso predictivo mayor al rendimiento académico. Esto insinúa que la deserción en la educación rural es un fenómeno multifactorial que requiere de modelos que integren datos situados para no incurrir en sesgos de generalidad urbana.

En tercer lugar, se concluye que predecir el abandono no es suficiente; es necesario explicarlo. El auge de técnicas como SHAP y LIME es importante para crear confianza en los resultados y permitir que los docentes y directivos comprendan los factores de riesgo particulares. Un sistema de alerta temprana solo es eficaz si el directivo o docente entiende el porqué de las razones detrás de las alertas, lo cual le facilita la toma de decisiones y el diseño de intervenciones pedagógicas.

Finalmente, la efectividad de estos modelos depende de su integración con sistemas nacionales de información, como el SIMAT y SIMPADE en Colombia. La sostenibilidad de estos sistemas de alerta temprana requiere que los datos se mantengan en tiempo real desde las instituciones educativas y que exista una articulación entre diferentes sectores (salud, trabajo, protección social) para abordar las diversas causas de la deserción de manera integral. Futuros estudios deben ahondar en el desarrollo de modelos con oportunidad cultural que incluyan variables concretas de las comunidades indígenas y campesinas.

En definitiva, el aprendizaje automático tiene el potencial de ser un motor de equidad en la educación, siempre que su diseño e implementación registren las particularidades estructurales de los territorios rurales. El éxito de estas iniciativas en la educación rural no debe medirse por lo complejo o sofisticado de los algoritmos, sino por su capacidad real para prevenir la deserción de los estudiantes en los territorios más sensibles.

REFERENCIAS

- Álamo, E. M.-C. (2024). Análisis de estrategias innovadoras para retención estudiantil con inteligencia artificial: una perspectiva multidisciplinaria. *European Public y Social Innovation Review*, 9, 1. <https://doi.org/10.31637/epsir-2024-440>
- Andrade-Girón, D., Sandivar-Rosas, J., Rodríguez, W. J. M., Susanibar-Ramirez, E., Toro-Dextre, E., Sánchez, J., Villarreal-Torres, H., y Ángeles-Morales, J. (2023). Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review [Review of Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review]. *ICST Transactions on Scalable Information Systems*. European Alliance for Innovation. <https://doi.org/10.4108/eetsis.3586>
- Bayona-Rodríguez, H., Naranjo Quintero, G. M., López Guarín, C. E., Rodríguez De Luque, J., Duque Márquez, I., Angulo González, M. V., Alarcón Párraga, C. L., Moreno Segura, N. E., Quiceno, S. I., Ortiz Vera, A. M., Vásquez Rojas, E., Quiceno Forero, S. I., Daza Malagón, C., Vásquez Rojas, E., Ortiz Vera, A. M., Aristizábal Henao, I. D., Bolívar Guzmán, E., Garavito Mariño, J., Herrera Aguilar, Z. T., ... Serrano Corredor, M. S. (2022). Deserción escolar en Colombia: análisis, determinantes y política de acogida, bienestar y permanencia.

- Bulut, O., Wongvorachan, T., He, S., y Lee, S. (2024a). Enhancing high-school dropout identification: a collaborative approach integrating human and machine insights. *Discover Education*, 3(1). <https://doi.org/10.1007/s44217-024-00209-4>
- Caballero, R. P. (2023). La tradición, la ruptura y la continuidad de la educación rural en el espacio iberoamericano. *Revista Boletín Redipe*, 12(2), 114. <https://doi.org/10.36260/rbr.v12i2.1937>
- Castillo-Peña, J. (2021). Expectativas y trayectorias educativas postsecundarias de jóvenes de territorios rurales en Chile. Una mirada desde el desarrollo humano. *Revista Iberoamericana de Educación Superior*, 127. <https://doi.org/10.22201/iisue.20072872e.2021.34.983>
- Castrillón-Gómez, O. D., Sarache, W., y Ruiz-Herrera, S. (2020). Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formación Universitaria*, 13(6), 217. <https://doi.org/10.4067/s0718-50062020000600217>
- Comisión Económica para América Latina y el Caribe (CEPAL), *Panorama Social de América Latina*, 2020 (LC/PUB.2021/2-P/Rev.1), Santiago, 2021.
- Cilleros, M. V. M., Sánchez-Prada, A., Álvarez, C. D., y Gómez, M. C. S. (2017). Valoración de un diseño educativo tecnológico para prevenir el abandono escolar. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 23, 61. <https://doi.org/10.17013/risti.23.61-77>
- Fitriana, S., Riniyanty, Laila, R., Pratama, S. A., y Lamasitudju, C. A. (2024). Prediksi Siswa Putus Sekolah Dan Keberhasilan Akademik Menggunakan Machine Learning. *Indonesian Journal of Computer Science*, 13(6). <https://doi.org/10.33022/ijcs.v13i6.4453>
- Forero-Corba, W., y Bennásar, F. N. (2023). Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: una revisión sistemática. *RIED Revista Iberoamericana de Educación a Distancia*, 27(1), 209. <https://doi.org/10.5944/ried.27.1.37491>
- Franco, J. C. L., Segura, Y. C. R., y Taborda, L. M. A. (2020). Diseños curriculares e inteligencia social en zonas urbanas y rurales de Colombia. *Cultura Educación y Sociedad*, 12(1), 255. <https://doi.org/10.17981/cultedusoc.12.1.2021.17>
- Galabay-Cajas, S. L., y Álvarez-Lozano, M. I. (2021). WhatsApp como estrategia educativa en pandemia: Una experiencia desde educación rural en Ecuador. *CIENCIAMATRIA*, 7(13), 397. <https://doi.org/10.35381/cm.v7i13.497>
- Galván Mora, L. (2020). Educación rural en América Latina: escenarios, tendencias y horizontes de investigación. <https://doi.org/10.24310/mgnmar.v1i2.8598>
- Gamboa-Cruzado, J., Alvarez-Cuellar, C. Y., Martinez-Medina, S., Chaparro, J. E. T., Damián, A. S., y Kong, M. P. R. (2023). Predicción de repitencias en estudiantes a nivel escolar usando Machine Learning: una revisión sistemática. *Apuntes Universitarios*, 13(2). <https://doi.org/10.17162/au.v13i2.1438>
- Guzmán, A., Moreno, S. P. B., y Vitery, F. C. (2021). Dropout in Rural Higher Education: A Systematic Review [Review of Dropout in Rural Higher Education: A Systematic Review]. *Frontiers in Education*, 6. *Frontiers Media*. <https://doi.org/10.3389/feduc.2021.727833>
- Haddaway, N. R., Collins, A. M., Coughlin, D., y Kirk, S. (2015). The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLOS ONE*, 10(9), 0. <https://doi.org/10.1371/journal.pone.0138237>
- Higgins, J. P. T., López-López, J. A., Becker, B. J., Davies, S. R., Dawson, S., Grimshaw, J. M., McGuinness, L. A., Moore, T. H. M., Rehfuss, E. A., Thomas, J., y Caldwell, D. M. (2019). Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Global Health*, 4, 0. <https://doi.org/10.1136/bmjgh-2018-000858>
- Hooshyar, D., Šir, G., Yang, Y., Kikas, E., Hämmäläinen, R., Kärkkäinen, T., Gašević, D., y Azevedo, R. (2025). Towards responsible AI for education: Hybrid human-AI to confront the Elephant in the room. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2504.16148>
- Huatangari, L. Q., Jara, D. M., Alvarado, N., Pino, M. E. M., y Gamarra, O. A. V. (2020). Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos. *Revista de Ciencia y Tecnología*, 34, 45. <https://doi.org/10.36995/j.recyt.2020.34.006>

- Jiménez-Rogel, J. P., y Campoverde-Moscol, A. I. (2024). Desafíos y perspectivas de la educación rural: acceso, permanencia y proyección profesional de los estudiantes. *MQRInvestigar*, 8(2), 3242. <https://doi.org/10.56048/mqr20225.8.2.2024.3242-3259>
- Khatun, Mst. R., Mim, M. A., Tasin, Md. M., y Hossain, Md. M. (2025). A hybrid framework of statistical, machine learning, and explainable AI methods for school dropout prediction. *PLoS ONE*, 20(9). <https://doi.org/10.1371/journal.pone.0331917>
- Leal, A. Z. (2019). Buenas prácticas pedagógicas y proyectos pedagógicos productivos : una experiencia en las instituciones educativas oficiales rurales del departamento del Valle del Cauca (Colombia). *Latinoamericana de Estudios Educativos*, 15(2), 11. <https://doi.org/10.17151/rlee.2019.15.2.2>
- Marcolino, M. S., Porto, T. R., Primo, T. T., Targino, R., Ramos, V. F. C., Queiroga, E. M., Muñoz, R., y Cechinel, C. (2025). Student dropout prediction through machine learning optimization: insights from moodle log data. *Scientific Reports*, 15(1), 9840. <https://doi.org/10.1038/s41598-025-93918-1>
- Mariñas, V. A. O., Orosco, L. S., y Yóplac, S. A. A. (2022). Brecha digital y educación virtual en instituciones educativas rurales. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 3(2), 534. <https://doi.org/10.56712/latam.v3i2.116>
- Melo, E. C., y Souza, F. S. H. de. (2023). Improving the prediction of school dropout with the support of the semi-supervised learning approach. *iSys - Brazilian Journal of Information Systems*, 16(1). <https://doi.org/10.5753/isys.2023.2852>
- Mínguez-Martínez, A., Sood, K., y Mahto, R. (2024). Early Detection of At-Risk Students Using Machine Learning. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.2412.09483>
- Montes, N., y Luna, F. (2024). Sistemas de alerta temprana: cómo acompañar trayectorias escolares inclusivas.
- Moons, K. G. M., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., Reitsma, J. B., y Collins, G. S. (2014). Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine*, 11(10), 0. <https://doi.org/10.1371/journal.pmed.1001744>
- Nagy, M., y Molontay, R. (2023). Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention. *International Journal of Artificial Intelligence in Education*, 34(2), 274. <https://doi.org/10.1007/s40593-023-00331-8>
- Niño, J. L. P., Guerrero, Ó. E. G., y Oliveros, D. C. L. (2024). Modelos de inteligencia artificial en minería de datos educativos para predecir la deserción en Educación Superior: una revisión integral. *Tecnura*, 28(82), 134. <https://doi.org/10.14483/22487638.23670>
- Ocen, S., Katalihwa, M., y Mwanje, D. (2025). Predicting Primary School Student Dropout Risk: A Machine Learning Framework for Early Intervention. *Journal of Intelligent Learning Systems and Applications*, 17(4), 267. <https://doi.org/10.4236/jilsa.2025.174017>
- Orozco, I. de la C., y Rubio, B. H. (2019). Asistencia y deserción escolar de la juventud indígena en Secundaria. *Revista Electrónica de Investigación Educativa*, 21, 1. <https://doi.org/10.24320/redie.2019.21.e24.1973>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J., Akl, E. A., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Parra-Sánchez, J. S., Torres, I. D., y Meriño, C. Y. M. de. (2023). Factores explicativos de la deserción universitaria abordados mediante inteligencia artificial. *Revista Electrónica de Investigación Educativa*, 25, 1. <https://doi.org/10.24320/redie.2023.25.e18.4455>
- Psyridou, M., Prezja, F., Torppa, M., Lerkkanen, M., Poikkeus, A., y Vasalampi, K. (2024). Machine learning predicts upper secondary education dropout as early as the end of primary school. *Scientific Reports*, 14(1), 12956. <https://doi.org/10.1038/s41598-024-63629-0>
- Quiñones Huatangari, L., y Carrasco Vega, Y. L. (2020). Rendimiento académico empleando minería de datos. <https://doi.org/10.48082/espacios-a20v41n44p17>
- Restrepo, M. G. (2022). Feature Reduction for Classification with Mixed Data : An Algorithmic Approach. *Research Portal Denmark*, 143. <https://local.forskningportal.dk/local/dki-cgi/ws/cris-link?src=cbsyid=cbs->

- 44a772b1-ff34-41c5-8ce9-f7a585a7e6d8yti=Feature%20Reduction%20for%20Classification%20with%20Mixed%20ata%20%3A%20An%20Algorithmic%20Approach
- Retamal, S. C., Casas, R. T., Hernández, A. E., y Clavería, K. A. (2021). Repensando la educación Rural.
- Rincón, A. G., Moreno, S. P. B., Cosenz, F., y Vitery, F. C. (2023). Prevention and Mitigation of Rural Higher Education Dropout in Colombia: A Dynamic Performance Management Approach. *F1000Research*, 12, 497. <https://doi.org/10.12688/f1000research.132267.1>
- Rincón, A. G., Moreno, S. P. B., Vitery, F. C., y Segovia-García, N. (2022). Deserción en la Educación Superior Rural: Análisis de Causas desde el Pensamiento Sistémico. *Qualitative Research in Education*, 11(2), 118. <https://doi.org/10.17583/qre.10048>
- Rodríguez, P., Villanueva, A., Dombrovskaja, L., y Valenzuela, J. P. (2023). A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: the case of Chile. *Education and Information Technologies*, 28(8), 10103. <https://doi.org/10.1007/s10639-022-11515-5>
- Roman, M. (2016). Factores Asociados al Abandono y la Deserción Escolar en América Latina: Una mirada de conjunto. *REICE Revista Iberoamericana Sobre Calidad Eficacia y Cambio En Educación*, 11(2). <https://doi.org/10.15366/reice2013.11.2.002>
- Romero, C., y Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3). <https://doi.org/10.1002/widm.1355>
- Salce, F. (2020). Deserción escolar y calidad de los docentes en Chile. *Revista de Análisis Económico*, 35(2), 135. <https://doi.org/10.4067/s0718-88702020000200135>
- Tete, M. F., Sousa, M. de M., Santana, T. S. D., y Silva, S. F. (2022). Aplicação de métodos preditivos em evasão no ensino superior: Uma revisão sistemática da literatura. *Education Policy Analysis Archives*, 30. <https://doi.org/10.14507/epaa.30.6845>
- Tong, T., y Li, Z. (2024). Predicting Learning Achievement Using Ensemble Learning with Result Explanation. *Research Square (Research Square)*. <https://doi.org/10.21203/rs.3.rs-4674228/v1>
- Torres, A. S. L. (2025). Desigualdad en la preparación académica de estudiantes rurales para la Educación Superior: Políticas de Inclusión Educativa en Ecuador. <https://doi.org/10.35537/10915/181437>
- UNESCO. (2021). Los sistemas de alerta temprana para prevenir el abandono escolar en América Latina y el Caribe.
- Villarrasa-Sapiña, I., García-Massó, X., Liébana, E., y Monfort-Torres, G. (2024). Academic achievement prediction in secondary education by decision tree analysis. *Educación XX1*, 27(1), 253. <https://doi.org/10.5944/educxx1.33351>
- Wolff, B., Mahoney, F., Lohiniva, A. L., y Corkum, M. (2019). Collecting and Analyzing Qualitative Data. *The CDC Field Epidemiology Manual*, 213. <https://doi.org/10.1093/oso/9780190933692.003.0010>
- Zambrano-Trujillo, S. L., Lara, F. L., y Cruz, Y. C. de la. (2022). Maestras de multigrado en Esmeraldas (Ecuador): una realidad silenciada. *Alteridad*, 17(2), 304. <https://doi.org/10.17163/alt.v17n2.2022.10>
- Zomaya, A. Y. (2016). Foreword. In *Elsevier eBooks*. Elsevier BV. <https://doi.org/10.1016/b978-0-12-804535-0.09995-0>

FUNDING / FINANCIAMIENTO

Not applicable / No aplica.

CONFLICT OF INTEREST / CONFLICTO DE INTERÉS

Not applicable / No aplica.

AUTHOR CONTRIBUTION / CONTRIBUCION AUTORAL

Conceptualization: Oscar Iván Montiel Petro

Data curation: Oscar Iván Montiel Petro

Formal analysis: Oscar Iván Montiel Petro

Research: Oscar Iván Montiel Petro

Methodology: Oscar Iván Montiel Petro

Validation: Oscar Iván Montiel Petro

Display: Oscar Iván Montiel Petro

Drafting - original draft: Oscar Iván Montiel Petro

Writing - proofreading and editing: Oscar Iván Montiel Petro