

Modelos de aprendizaje automático para la detección temprana de factores de riesgo en el desempeño académico de estudiantes universitarios mediante variables socio-tecnológicas

Machine learning models for the early detection of risk factors in the academic performance of university students using socio-technological variables

José Fabián Cevallos Mazón; Walter Fabian Taday Guashpa; Tatiana Paola Sánchez Gavilanes; Wilter Alexander Camacho Yunda; Carlos Santiago Ramos Vaca; Teresita Lourdes Argüello Estrella

Cita:

Cevallos Mazón, J. F., et al. (2026). Modelos de aprendizaje automático para la detección temprana de factores de riesgo en el desempeño académico de estudiantes universitarios mediante variables socio-tecnológicas. *TESLA Revista Científica*, 6(2), e754.

<https://doi.org/10.55204/trc.v6i2.e754>

Recibido: 25/09/2026

Revisado: 08/10/2026

Corregido: 14/10/2026

Aceptado: 19/10/2026

Publicado: 20/10/2026

Licencia:

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.

1 Universidad Nacional de Chimborazo UNACH, 2 Escuela Superior Politécnica de Chimborazo ESPOCH, 3-4 Universidad Estatal de Bolívar UEB, 5-6 Investigador Externo

1 josef.cevallos@unach.edu.ec, 2 walter.taday@epoch.edu.ec, 3 tsanchez@ueb.edu.ec, 4 wicamacho@ueb.edu.ec, 5 psicologiasantyramos2023@gmail.com, 6 teresita.arguello@docentes.educacion.edu.ec

ORCID: 1: 0009-0009-4703-2708; 2: 0009-0009-6801-6248; 3: 0009-0009-4063-9160; 4: 0000-0002-6176-2864; 5: 0009-0004-6931-5163; 6: 0009-0002-1428-6719

Resumen: La presente investigación analiza los modelos de aprendizaje automático aplicados a la detección temprana de factores de riesgo en el desempeño académico de estudiantes universitarios, con énfasis en la integración de variables socio-tecnológicas derivadas de plataformas de gestión del aprendizaje (LMS). Se llevó a cabo una revisión sistemática de literatura siguiendo el protocolo PRISMA 2020, en la cual se analizaron 40 estudios publicados entre 2020 y 2026 provenientes de bases de datos internacionales y latinoamericanas. Los hallazgos principales indican que los métodos de ensamble, tales como Random Forest, XGBoost y Gradient Boosting, constituyen el estado del arte en predicción del riesgo académico, con valores de exactitud entre el 85 % y el 93 % y un área bajo la curva (AUC-ROC) superior a 0,90. Se determinó además que la incorporación de variables de comportamiento digital frecuencia de acceso al LMS, tiempo de sesión, participación en foros y puntualidad en entregas incrementa el poder predictivo de los modelos entre 7 y 14 puntos porcentuales frente al uso exclusivo de calificaciones históricas. Se propone una arquitectura híbrida que integre modelos predictivos supervisados con agentes de aprendizaje por refuerzo para la generación de intervenciones pedagógicas personalizadas en tiempo real dentro de la infraestructura institucional.

Palabras clave: aprendizaje automático, riesgo académico, variables socio-tecnológicas, detección temprana, LMS, métodos de ensamble, PRISMA.

Abstract: This study examines machine learning models applied to the early detection of risk factors in university students' academic performance, with particular emphasis on integrating socio-technological variables derived from Learning Management System (LMS) platforms. A systematic literature review was conducted following the PRISMA 2020 protocol, analyzing 40 studies published between 2020 and 2026 retrieved from international and Latin American databases. The main findings indicate that ensemble methods, namely Random Forest, XGBoost, and Gradient Boosting, represent the current state of the art in academic risk prediction, achieving accuracy values between 85 % and 93 % with an area under the curve (AUC-ROC) exceeding 0.90. Furthermore, the incorporation of digital behavior variables — LMS access frequency, session duration, forum participation, and assignment submission punctuality was found to increase the predictive power of models by 7 to 14 percentage points compared to the exclusive use of historical grades. A hybrid architecture integrating supervised predictive models with reinforcement learning agents is proposed for generating personalized pedagogical interventions in real time within institutional infrastructure.

Keywords: machine learning; academic risk; socio-technological variables; early detection; LMS; ensemble methods; PRISMA.

INTRODUCCIÓN

La deserción universitaria en carreras de ingeniería y tecnologías de la información en América Latina constituye un problema estructural de primer orden, con tasas que superan el 35 % durante los dos primeros años

de formación (Jiménez et al., 2025; Rubio-Tobar, 2025; Arroyo Saltos et al., 2025). Este fenómeno genera consecuencias sociales y económicas de alto impacto, al truncar trayectorias educativas que representan inversiones significativas tanto para las familias como para los Estados. En el contexto ecuatoriano, factores diferenciados como la brecha de conectividad en zonas rurales, la condición laboral activa de una proporción considerable de estudiantes y la transición abrupta hacia la virtualidad postpandemia agravan este escenario de forma particular (Cajahuanca et al., 2022; Segovia-García et al., 2024; Del Carpio-Mendoza, 2024).

El bajo rendimiento académico no se manifiesta como un evento repentino, sino como la acumulación progresiva de señales de alerta que, al no ser detectadas a tiempo, derivan en la imposibilidad de revertir el deterioro (Jimenez-Garcia et al., 2025; Benitez Mero et al., 2026). Este carácter multidimensional del fenómeno exige herramientas analíticas capaces de procesar simultáneamente variables de distinta naturaleza: académicas, socioeconómicas, comportamentales y tecnológicas (Gusqui & Alejandro, 2022; Pinzon Aparicio, 2024). Los enfoques estadísticos tradicionales resultan insuficientes ante esta complejidad, dado que no logran capturar las interacciones no lineales entre los predictores del riesgo.

El aprendizaje automático machine learning se ha consolidado como el paradigma metodológico de mayor eficacia para la predicción del riesgo académico a partir de datos institucionales heterogéneos. La revisión sistemática de Namoun y Alshanqiti (2021), que sintetiza más de un centenar de estudios, confirma que los algoritmos de clasificación supervisada superan consistentemente a los métodos estadísticos clásicos. Esta evidencia se corrobora en las revisiones independientes de Balaji et al. (2021), Baashar et al. (2021) y Del Carpio-Mendoza (2024), que identifican a los métodos de ensamble como la familia de mayor rendimiento predictivo en contextos educativos.

Desde una perspectiva algorítmica, los modelos aplicados al rendimiento estudiantil se organizan en cuatro familias principales. La primera comprende los clasificadores clásicos árboles de decisión, Naive Bayes, regresión logística y K-Nearest Neighbors, caracterizados por su alta interpretabilidad y bajo costo computacional, con valores de exactitud reportados entre el 72 % y el 82 % (Rico Páez & Gaytan Ramírez, 2022; Rico Páez, 2022; Parraga Menchaca, 2024; Páez & Ramírez, 2022). La segunda familia corresponde a los métodos de ensamble, que constituyen el estado del arte (Yagci, 2022; Adnan et al., 2021; Chen & Zhai, 2023; Contreas-Bravo et al., 2023; Abdrakhmanov et al., 2024).

Los estudios comparativos de Yagci (2022) y Chen y Zhai (2023) sitúan de manera consistente a Random Forest y XGBoost como los algoritmos líderes en métricas de exactitud (85-93 %), precisión y AUC-ROC (>0,90) sobre conjuntos de datos educativos de composición heterogénea. La tercera familia, las redes neuronales artificiales superficiales, ofrece capacidad de modelización no lineal con rendimiento comparable al de los ensambles cuando se dispone de conjuntos de datos de tamaño intermedio (Namoun & Alshanqiti, 2021; Balaji et al., 2021). La cuarta familia, los modelos de aprendizaje profundo, lidera en escenarios con datos secuenciales longitudinales (Xu et al., 2023; Al-Ali et al., 2024).

La proliferación de plataformas de gestión del aprendizaje (LMS) como Moodle y Canvas ha generado un volumen de datos de comportamiento digital sin precedentes: frecuencia y duración de sesiones, acceso a materiales, participación en foros y puntualidad en entregas (Raza et al., 2021; Achuthan et al., 2021; Schurmann & Quaiser-Pohl, 2022). Estas variables, denominadas socio-tecnológicas, actúan como señales tempranas del compromiso cognitivo y la autorregulación del aprendizaje, disponibles desde las primeras semanas del semestre. Al-Ali et al. (2024) demuestran que su incorporación incrementa el poder predictivo entre 7 y 14 puntos porcentuales frente al uso exclusivo de calificaciones históricas.

La selección de características constituye un eslabón crítico en el pipeline de ML educativo. Hussain et al. (2021) proponen un algoritmo híbrido de optimización metaheurística que supera a los métodos clásicos de selección en escenarios de alta dimensionalidad. Zaffar et al. (2021) documentan mejoras de hasta 8 puntos porcentuales en la métrica F1 al aplicar selección basada en importancia de Gini antes del entrenamiento del modelo. La toma de decisiones multicriterio para la configuración de pipelines, abordada por Liu, H.-C. et al. (2022) mediante lógica difusa, complementa este marco con una capa de gobernanza. Thai-Nghe y Schmidt-Thieme (2021) demuestran que las máquinas de factorización capturan relaciones latentes en entornos colaborativos.

En el horizonte emergente, la integración del aprendizaje por refuerzo (RL) con modelos supervisados representa el avance de mayor impacto potencial. El trabajo fundacional de Afsar et al. (2022) documenta la capacidad de los agentes de RL para personalizar intervenciones de forma dinámica en sistemas de recomendación educativos. Molenaar (2022) conceptualiza la regulación híbrida humano-IA, donde el docente retiene la agencia pedagógica de alto nivel. Suin Guerrero et al. (2024) adaptan este enfoque al contexto latinoamericano. Los fundamentos teóricos del RL se documentan en Miyamoto et al. (2021), Pérez (2024) e IBM (2024), con aplicaciones en modelos de relaciones complejas (Liu, X. et al., 2021).

El presente trabajo se propone sistematizar el estado del arte de los modelos de ML para la detección temprana del riesgo académico universitario, con énfasis en variables socio-tecnológicas de plataformas LMS, mediante una revisión sistemática bajo protocolo PRISMA 2020 (Page et al., 2021; Rethlefsen et al., 2021). La hipótesis directriz postula que la combinación de variables de comportamiento digital, indicadores socioeconómicos y registros históricos de rendimiento, procesados mediante métodos de ensamble o arquitecturas profundas, permite predecir el riesgo académico con exactitud superior al 85 %, habilitando intervenciones pedagógicas oportunas en universidades latinoamericanas.

Contexto general del problema

Importancia de la educación superior en la sociedad

La educación superior desempeña un papel fundamental en el desarrollo social, económico y tecnológico de las naciones, al constituirse como un pilar esencial para la formación de capital humano altamente calificado. Las instituciones universitarias no solo contribuyen a la generación y transmisión del conocimiento, sino que también impulsan la innovación, la investigación científica y el progreso sostenible. En este sentido, el acceso, permanencia y éxito en la educación superior son considerados factores clave para el crecimiento de los países y la reducción de desigualdades sociales (UNESCO, 2021).

Asimismo, la educación superior influye directamente en la empleabilidad de los individuos y en su capacidad para adaptarse a entornos laborales cada vez más dinámicos y tecnológicos. Por ello, garantizar un adecuado desempeño académico durante la formación universitaria no solo beneficia al estudiante a nivel individual, sino que también repercute positivamente en la sociedad en su conjunto (Banco Mundial, 2020).

Problemática del bajo rendimiento académico

A pesar de la importancia de la educación superior, el bajo rendimiento académico continúa siendo una problemática persistente en universidades a nivel global. Este fenómeno se manifiesta a través de bajas calificaciones, repetición de asignaturas y dificultades en el proceso de aprendizaje, lo que puede afectar significativamente la trayectoria educativa de los estudiantes.

Diversas investigaciones han evidenciado que el bajo rendimiento académico está asociado con múltiples factores que van más allá de las capacidades cognitivas, incluyendo aspectos emocionales, sociales y contextuales. Además, un desempeño académico deficiente puede generar desmotivación, estrés y una disminución en el compromiso del estudiante con su proceso educativo (García-Martínez et al., 2021).

Impacto de la deserción estudiantil

Uno de los principales efectos del bajo rendimiento académico es la deserción estudiantil, considerada uno de los mayores desafíos de la educación superior en la actualidad. La deserción no solo implica la interrupción de los estudios por parte del estudiante, sino que también representa una pérdida significativa de recursos para las instituciones educativas y la sociedad.

En América Latina, las tasas de deserción son particularmente preocupantes, ya que un número considerable de estudiantes no logra completar sus estudios universitarios. Este fenómeno está influenciado por factores

económicos, sociales y académicos, lo que evidencia la necesidad de abordar el problema desde una perspectiva integral (Banco Mundial, 2020).

Necesidad de mejorar la calidad educativa

Ante este panorama, la mejora de la calidad educativa se ha convertido en una prioridad para las instituciones de educación superior. Esto implica no solo optimizar los procesos de enseñanza y aprendizaje, sino también desarrollar mecanismos que permitan identificar y atender las dificultades académicas de los estudiantes de manera oportuna.

En este sentido, la detección temprana de factores de riesgo en el desempeño académico es clave para implementar intervenciones efectivas que contribuyan a mejorar los resultados educativos. La incorporación de enfoques innovadores, como el uso de tecnologías digitales y el análisis de datos, permite a las instituciones adoptar estrategias más eficientes y basadas en evidencia (Ifenthaler & Yau, 2020).

Asimismo, la integración de herramientas tecnológicas en el ámbito educativo ha transformado la forma en que se enseña y se aprende, generando nuevas oportunidades, pero también nuevos desafíos. Por ello, resulta imprescindible comprender el contexto actual de la educación superior y las variables que influyen en el rendimiento académico, con el fin de diseñar soluciones que respondan a las necesidades reales de los estudiantes.

Complejidad del desempeño académico

Naturaleza multifactorial del rendimiento académico

El rendimiento académico en la educación superior es un fenómeno complejo que no puede ser explicado únicamente a partir de las capacidades cognitivas del estudiante. Diversas investigaciones coinciden en que este depende de una combinación de factores individuales, sociales, económicos y tecnológicos que interactúan de manera dinámica y no lineal (Alhassan et al., 2022; García-Martínez et al., 2021). Esta naturaleza multifactorial implica que el desempeño académico debe ser analizado desde un enfoque integral, considerando múltiples dimensiones que influyen en el proceso de aprendizaje.

En este sentido, estudios recientes han demostrado que variables como la motivación, la autorregulación del aprendizaje, el contexto socioeconómico y el acceso a recursos educativos tienen un impacto significativo en los resultados académicos (Broadbent & Poon, 2021). Asimismo, la diversidad de perfiles estudiantiles en la educación superior refuerza la necesidad de comprender el rendimiento académico como un fenómeno heterogéneo, donde cada estudiante presenta características y condiciones particulares.

Adicionalmente, el carácter multifactorial del rendimiento académico plantea desafíos importantes para su análisis y predicción, ya que requiere la integración de grandes volúmenes de datos provenientes de diferentes fuentes. En este contexto, el uso de enfoques analíticos avanzados permite identificar patrones complejos y relaciones entre variables que no son evidentes mediante métodos tradicionales (Ifenthaler & Yau, 2020).

Interacción entre factores individuales, sociales y tecnológicos

El desempeño académico es el resultado de la interacción entre múltiples factores que actúan de forma simultánea. Entre los factores individuales se incluyen aspectos como las habilidades cognitivas, la motivación, la disciplina y las estrategias de aprendizaje. Estos elementos influyen directamente en la capacidad del estudiante para adquirir conocimientos y adaptarse a las exigencias del entorno académico (Richardson et al., 2021).

Por otro lado, los factores sociales, como el entorno familiar, el nivel socioeconómico y el apoyo social, desempeñan un papel fundamental en el rendimiento académico. Estudiantes provenientes de contextos

socioeconómicos desfavorecidos suelen enfrentar mayores dificultades debido a limitaciones en el acceso a recursos educativos, lo que puede afectar su desempeño (OECD, 2020). Asimismo, el apoyo familiar y el entorno social influyen en la motivación y el compromiso del estudiante con sus estudios.

En los últimos años, los factores tecnológicos han adquirido una relevancia significativa, especialmente con la expansión de la educación digital y el aprendizaje en línea. El acceso a internet, el uso de plataformas virtuales y la disponibilidad de herramientas digitales se han convertido en elementos clave para el éxito académico (Bond et al., 2021). Sin embargo, la existencia de una brecha digital puede generar desigualdades en el acceso a estas tecnologías, afectando negativamente el rendimiento de ciertos grupos de estudiantes.

La interacción entre estos factores no es lineal, sino que se caracteriza por relaciones complejas que pueden potenciar o limitar el desempeño académico. Por ejemplo, un estudiante con alta motivación puede compensar ciertas limitaciones tecnológicas, mientras que la falta de apoyo social puede afectar negativamente incluso a estudiantes con buenas capacidades cognitivas. Esta interdependencia refuerza la necesidad de adoptar enfoques integrales para el análisis del rendimiento académico (Zawacki-Richter et al., 2020).

Factores que influyen en el desempeño académico

Factores sociales

Los factores sociales constituyen una de las dimensiones más influyentes en el rendimiento académico, ya que están relacionados con el contexto en el que se desarrolla el estudiante. Entre estos factores, el nivel socioeconómico ha sido ampliamente estudiado, evidenciando que estudiantes provenientes de contextos con mayores recursos tienden a presentar mejores resultados académicos debido al acceso a materiales educativos, apoyo adicional y condiciones adecuadas de estudio (OECD, 2020).

Asimismo, el entorno familiar desempeña un papel clave en el desarrollo académico. El apoyo de los padres, la estabilidad familiar y el nivel educativo del hogar influyen directamente en la motivación, disciplina y hábitos de estudio del estudiante (Sirin, 2021). En este sentido, un entorno familiar favorable puede potenciar el rendimiento académico, mientras que condiciones adversas pueden dificultar el proceso de aprendizaje.

Por otro lado, aspectos como la carga laboral, las responsabilidades familiares y el contexto social también pueden afectar el desempeño académico, especialmente en estudiantes universitarios que deben equilibrar múltiples responsabilidades. Estas condiciones pueden generar estrés, fatiga y limitaciones en el tiempo disponible para el estudio, impactando negativamente en los resultados académicos (Alhassan et al., 2022).

Factores académicos

Los factores académicos están directamente relacionados con el proceso de enseñanza-aprendizaje y con las prácticas educativas dentro de las instituciones. Entre estos, los hábitos de estudio son uno de los elementos más determinantes del rendimiento académico. La planificación del tiempo, la constancia y el uso de estrategias de aprendizaje adecuadas permiten a los estudiantes mejorar su desempeño (Broadbent & Poon, 2021).

La asistencia a clases y la participación también son variables relevantes, ya que facilitan la comprensión de los contenidos y el desarrollo de habilidades académicas. Estudios recientes han demostrado que los estudiantes que participan activamente en su proceso de aprendizaje tienden a obtener mejores resultados (Richardson et al., 2021).

Asimismo, los métodos de enseñanza utilizados por los docentes influyen significativamente en el rendimiento académico. Estrategias pedagógicas innovadoras, como el aprendizaje activo y el uso de tecnologías educativas, pueden mejorar la comprensión de los contenidos y aumentar el compromiso del estudiante (Bond et al., 2021).

Finalmente, los sistemas de evaluación también juegan un papel importante, ya que no solo miden el desempeño académico, sino que también influyen en la forma en que los estudiantes abordan su aprendizaje.

Evaluaciones formativas y continuas pueden contribuir a una mejor comprensión de los contenidos y a un seguimiento más efectivo del progreso académico.

Factores tecnológicos

En la actualidad, los factores tecnológicos han adquirido una relevancia creciente en el desempeño académico, especialmente debido a la expansión de la educación digital. El acceso a internet es uno de los principales determinantes, ya que permite a los estudiantes acceder a recursos educativos, plataformas virtuales y herramientas de aprendizaje en línea (Bond et al., 2021).

El uso de plataformas educativas virtuales, como sistemas de gestión del aprendizaje (LMS), facilita la interacción entre estudiantes y docentes, así como el acceso a contenidos académicos. Sin embargo, el aprovechamiento de estas herramientas depende de las competencias digitales de los estudiantes y de su capacidad para utilizarlas de manera efectiva (Ifenthaler & Yau, 2020).

Por otro lado, la disponibilidad de dispositivos tecnológicos, como computadoras y smartphones, influye en la capacidad del estudiante para participar en entornos de aprendizaje digital. La falta de acceso a estos recursos puede generar desigualdades en el rendimiento académico, especialmente en contextos socioeconómicos desfavorecidos.

Problemática de la detección tardía del riesgo académico

La detección tardía del riesgo académico constituye una de las principales limitaciones en los sistemas educativos actuales, especialmente en el nivel superior. A pesar de los avances en la gestión educativa, muchas instituciones continúan identificando a los estudiantes en riesgo cuando ya presentan un bajo rendimiento consolidado, lo que reduce significativamente la efectividad de las intervenciones (García-Martínez et al., 2021). Esta situación pone en evidencia la necesidad de adoptar enfoques más proactivos que permitan anticipar problemas académicos antes de que se agraven.

Falta de identificación temprana de estudiantes en riesgo

Uno de los principales problemas en la educación superior es la ausencia de mecanismos eficientes para identificar de manera temprana a los estudiantes que presentan riesgo académico. En muchos casos, la detección se realiza a partir de resultados finales, como calificaciones bajas o reprobación de asignaturas, lo que implica que el estudiante ya ha experimentado dificultades significativas en su proceso de aprendizaje (Ifenthaler & Yau, 2020).

Esta falta de identificación temprana limita la capacidad de las instituciones para implementar estrategias preventivas. Estudios recientes destacan que la intervención oportuna puede mejorar considerablemente el rendimiento académico y reducir la probabilidad de deserción, lo que resalta la importancia de contar con sistemas que permitan detectar señales de alerta en etapas iniciales (Tempelaar et al., 2020).

Limitaciones de los métodos tradicionales de evaluación

Los métodos tradicionales de evaluación, basados principalmente en exámenes y calificaciones finales, presentan limitaciones importantes para la detección del riesgo académico. Estos enfoques suelen centrarse en medir resultados, pero no permiten comprender el proceso de aprendizaje ni identificar las causas subyacentes de un bajo rendimiento (Zawacki-Richter et al., 2020).

Además, estos métodos suelen ser reactivos en lugar de preventivos, ya que se aplican después de que el estudiante ha enfrentado dificultades. Esto dificulta la implementación de acciones correctivas a tiempo y reduce las posibilidades de mejorar el desempeño académico. En este sentido, la dependencia exclusiva

de evaluaciones tradicionales puede generar una visión incompleta del estudiante y de sus necesidades educativas (García-Martínez et al., 2021).

Necesidad de soluciones basadas en datos

Importancia del análisis de datos en la educación

El análisis de datos en educación ha cobrado relevancia en los últimos años debido a su capacidad para transformar información en conocimiento útil. A través del procesamiento de datos académicos, es posible identificar tendencias, evaluar el desempeño estudiantil y detectar señales tempranas de riesgo académico (García-Martínez et al., 2021).

Además, el análisis de datos permite integrar información proveniente de múltiples fuentes, como registros académicos, interacción en plataformas virtuales y variables socioeconómicas, lo que facilita una visión más completa del estudiante. Este enfoque integral contribuye a mejorar la comprensión del proceso de aprendizaje y a diseñar estrategias más efectivas para el éxito académico (Tempelaar et al., 2020).

Asimismo, el uso de técnicas analíticas avanzadas permite descubrir relaciones complejas entre variables que no son evidentes mediante métodos convencionales, lo que representa una ventaja significativa en el estudio del rendimiento académico.

Toma de decisiones basada en evidencia

Uno de los principales beneficios de las soluciones basadas en datos es la posibilidad de apoyar la toma de decisiones en evidencia empírica. En el ámbito educativo, esto implica utilizar información objetiva para diseñar políticas, estrategias pedagógicas e intervenciones orientadas a mejorar el rendimiento académico (Siemens & Baker, 2021).

La toma de decisiones basada en datos permite a las instituciones identificar de manera precisa a los estudiantes que requieren apoyo, optimizar la asignación de recursos y evaluar la efectividad de las acciones implementadas. De esta manera, se reduce la incertidumbre en la gestión educativa y se incrementa la probabilidad de obtener resultados positivos.

Por otro lado, este enfoque también favorece la personalización del aprendizaje, ya que permite adaptar las estrategias educativas a las necesidades específicas de cada estudiante, considerando sus características y condiciones particulares (Ifenthaler & Yau, 2020).

Evolución hacia la analítica del aprendizaje

La analítica del aprendizaje (Learning Analytics) ha emergido como un campo clave en la aplicación de soluciones basadas en datos en educación. Este enfoque se centra en la recopilación, análisis e interpretación de datos relacionados con el aprendizaje, con el objetivo de mejorar los procesos educativos y los resultados académicos (Siemens & Baker, 2021).

A través de la analítica del aprendizaje, es posible monitorear el comportamiento de los estudiantes en tiempo real, identificar patrones de participación y detectar posibles dificultades en el proceso de aprendizaje. Esto permite generar alertas tempranas y diseñar intervenciones oportunas que contribuyan a mejorar el rendimiento académico (Ifenthaler & Yau, 2020).

Además, la analítica del aprendizaje facilita la evaluación continua del progreso estudiantil, lo que permite realizar ajustes en las estrategias pedagógicas de manera dinámica. Este enfoque representa un cambio significativo respecto a los métodos tradicionales, ya que se basa en un análisis constante y sistemático de la información.

Introducción al aprendizaje automático

Concepto y fundamentos del aprendizaje automático

El aprendizaje automático puede definirse como una disciplina que desarrolla modelos computacionales capaces de aprender patrones a partir de datos y realizar predicciones o tomar decisiones sin intervención humana directa. Este proceso se basa en el entrenamiento de algoritmos mediante conjuntos de datos, a partir de los cuales el modelo identifica relaciones entre variables (Alhassan et al., 2022).

El funcionamiento del aprendizaje automático implica varias etapas, entre las que se incluyen la recopilación de datos, el preprocesamiento, la selección de características, el entrenamiento del modelo y la evaluación de su desempeño. Cada una de estas etapas es fundamental para garantizar la calidad de los resultados obtenidos y la capacidad del modelo para generalizar a nuevos datos (Goodfellow et al., 2021).

Además, el aprendizaje automático se apoya en conceptos estadísticos y matemáticos que permiten modelar la incertidumbre y evaluar la precisión de las predicciones. Esto lo convierte en una herramienta robusta para el análisis de datos complejos, como los que se generan en entornos educativos.

Tipos de aprendizaje automático

El aprendizaje automático puede clasificarse en diferentes tipos según la forma en que se utilizan los datos durante el proceso de entrenamiento. Entre los principales enfoques se encuentran el aprendizaje supervisado, no supervisado y por refuerzo.

El aprendizaje supervisado es uno de los más utilizados en el ámbito educativo, ya que se basa en el uso de datos etiquetados para entrenar modelos que puedan predecir resultados específicos, como el rendimiento académico o la probabilidad de deserción (García-Martínez et al., 2021). En este enfoque, el modelo aprende a partir de ejemplos en los que se conoce la relación entre las variables de entrada y la salida.

Por otro lado, el aprendizaje no supervisado se utiliza para identificar patrones o estructuras en los datos sin necesidad de etiquetas. Este enfoque es útil para segmentar estudiantes en grupos con características similares o para descubrir relaciones ocultas entre variables (Zawacki-Richter et al., 2020).

Finalmente, el aprendizaje por refuerzo se basa en la interacción del modelo con un entorno, donde aprende a tomar decisiones a partir de recompensas o penalizaciones. Aunque su aplicación en educación es menos común, tiene potencial en sistemas adaptativos de aprendizaje (Alhassan et al., 2022).

Aplicación del aprendizaje automático en educación

El aprendizaje automático ha transformado significativamente el ámbito educativo al proporcionar herramientas avanzadas para el análisis de datos y la mejora de los procesos de enseñanza-aprendizaje. En particular, su aplicación en la educación superior ha permitido abordar problemáticas complejas como la predicción del rendimiento académico, la identificación de estudiantes en riesgo y la personalización del aprendizaje (Zawacki-Richter et al., 2020).

A través del uso de modelos predictivos, las instituciones educativas pueden aprovechar los datos generados por los estudiantes para obtener información valiosa que contribuya a la toma de decisiones informadas y al diseño de estrategias educativas más efectivas (García-Martínez et al., 2021).

Predicción del rendimiento académico

Una de las principales aplicaciones del aprendizaje automático en educación es la predicción del rendimiento académico. Mediante el análisis de datos históricos, como calificaciones, asistencia, interacción en plataformas virtuales y variables socioeconómicas, los modelos de aprendizaje automático pueden estimar el desempeño futuro de los estudiantes (Ahmed et al., 2021).

Estos modelos permiten identificar patrones que están asociados con el éxito o el fracaso académico, lo que facilita la detección de estudiantes que podrían presentar dificultades en su proceso de aprendizaje. De esta manera, las instituciones pueden anticiparse a posibles problemas y tomar medidas preventivas que contribuyan a mejorar los resultados académicos (García-Martínez et al., 2021).

Además, la predicción del rendimiento académico no solo permite clasificar a los estudiantes según su nivel de riesgo, sino también comprender los factores que influyen en su desempeño, lo que resulta fundamental para el diseño de estrategias de intervención.

Identificación de estudiantes en riesgo

Otra aplicación clave del aprendizaje automático es la identificación temprana de estudiantes en riesgo académico. A través del análisis de múltiples variables, estos modelos pueden detectar señales de alerta antes de que se reflejen en calificaciones bajas o en la deserción estudiantil (Ifenthaler & Yau, 2020).

Por ejemplo, patrones como la disminución en la participación en plataformas virtuales, la baja frecuencia de acceso a materiales educativos o el retraso en la entrega de tareas pueden ser indicadores de riesgo académico. El aprendizaje automático permite analizar estos patrones de manera automatizada y generar alertas tempranas que faciliten la intervención oportuna (Tempelaar et al., 2020).

La identificación temprana de estudiantes en riesgo es fundamental para mejorar la retención estudiantil y reducir la deserción, ya que permite implementar acciones correctivas en etapas iniciales del proceso educativo.

Personalización del aprendizaje

El aprendizaje automático también permite avanzar hacia modelos educativos más personalizados, adaptados a las necesidades y características de cada estudiante. A partir del análisis de datos individuales, es posible diseñar experiencias de aprendizaje que se ajusten al ritmo, estilo y nivel de conocimiento del estudiante (Bond et al., 2021).

La personalización del aprendizaje contribuye a mejorar la motivación, el compromiso y el rendimiento académico, ya que los estudiantes reciben contenidos y actividades acordes a sus necesidades específicas. Además, permite a los docentes identificar áreas de mejora y brindar apoyo de manera más efectiva.

En este contexto, los sistemas adaptativos de aprendizaje, basados en algoritmos de aprendizaje automático, representan una herramienta innovadora para mejorar la calidad educativa y promover un aprendizaje más eficiente (Zawacki-Richter et al., 2020).

Casos de uso en educación superior

El uso del aprendizaje automático en la educación superior se ha incrementado en los últimos años, con múltiples aplicaciones en universidades a nivel global. Entre los casos de uso más relevantes se encuentran los sistemas de alerta temprana, que permiten identificar estudiantes en riesgo, y los modelos predictivos de rendimiento académico, que facilitan la toma de decisiones institucionales (García-Martínez et al., 2021).

Asimismo, algunas instituciones han implementado sistemas de recomendación académica que sugieren cursos, recursos educativos y estrategias de estudio personalizadas, en función del perfil del estudiante. Estas herramientas contribuyen a mejorar la experiencia educativa y a optimizar los procesos de aprendizaje.

Por otro lado, el aprendizaje automático también se utiliza en la evaluación del desempeño docente, el análisis de la calidad educativa y la optimización de la gestión institucional. Estas aplicaciones demuestran el potencial de esta tecnología para transformar el sistema educativo y responder a los desafíos actuales de la educación superior (Ifenthaler & Yau, 2020).

Importancia de las variables socio-tecnológicas en el rendimiento académico

Integración de dimensiones sociales y tecnológicas

Las variables socio-tecnológicas integran factores relacionados con el contexto social del estudiante y su interacción con las tecnologías digitales. Entre las variables sociales se incluyen el nivel socioeconómico, el entorno familiar, el acceso a recursos educativos y las condiciones de vida, mientras que las variables tecnológicas abarcan el acceso a internet, el uso de plataformas virtuales y la disponibilidad de dispositivos tecnológicos (OECD, 2020; Bond et al., 2021).

La integración de estas dimensiones permite analizar cómo las condiciones sociales influyen en el acceso y uso de la tecnología, y cómo estos factores, a su vez, impactan en el rendimiento académico. Por ejemplo, estudiantes con acceso limitado a internet pueden enfrentar dificultades para participar en actividades educativas en línea, lo que afecta su desempeño.

Asimismo, la interacción entre variables sociales y tecnológicas puede generar desigualdades en el aprendizaje, especialmente en contextos donde la brecha digital es significativa. Esta situación resalta la importancia de considerar ambas dimensiones de manera conjunta en el análisis del rendimiento académico.

Relevancia de las variables socio-tecnológicas en modelos predictivos

La incorporación de variables socio-tecnológicas en modelos de aprendizaje automático permite mejorar la capacidad predictiva de estos sistemas. Estudios recientes han demostrado que los modelos que incluyen información sobre el contexto social y el uso de tecnología presentan mayor precisión en la predicción del rendimiento académico en comparación con aquellos que utilizan únicamente variables académicas (García-Martínez et al., 2021).

Estas variables permiten identificar patrones más complejos y relaciones no evidentes entre los factores que influyen en el aprendizaje. Por ejemplo, el análisis conjunto de la participación en plataformas virtuales y el nivel socioeconómico puede revelar diferencias significativas en el comportamiento académico de los estudiantes (Ifenthaler & Yau, 2020).

Además, la inclusión de variables socio-tecnológicas facilita la identificación de estudiantes en situaciones de vulnerabilidad, lo que permite diseñar estrategias de intervención más efectivas y adaptadas a sus necesidades específicas.

Relación entre variables socio-tecnológicas y rendimiento académico

Diversas investigaciones han evidenciado la relación directa entre las variables socio-tecnológicas y el rendimiento académico. El acceso a recursos tecnológicos, como computadoras y conexión a internet, ha demostrado ser un factor determinante en el éxito académico, especialmente en entornos de educación virtual o híbrida (Bond et al., 2021).

Por otro lado, el nivel socioeconómico influye en la disponibilidad de estos recursos y en las condiciones de estudio, lo que puede generar desigualdades en el rendimiento académico. Estudiantes con mayores recursos tienden a tener mejores condiciones para el aprendizaje, mientras que aquellos en contextos desfavorecidos enfrentan mayores dificultades (OECD, 2020).

Asimismo, el uso adecuado de herramientas digitales puede potenciar el aprendizaje, facilitando el acceso a información, la interacción con docentes y la participación en actividades académicas. Sin embargo, la falta de competencias digitales puede limitar el aprovechamiento de estas herramientas, afectando negativamente el desempeño académico (Ifenthaler & Yau, 2020).

METODOLOGIA

2.1. Pregunta de investigación y criterios de elegibilidad

La revisión sistemática se estructuró en torno a una pregunta de investigación formulada bajo el marco PICO (Population, Intervention, Comparator, Outcome), estándar recomendado por PRISMA 2020 (Page et al., 2021) para garantizar precisión y reproducibilidad. Los componentes se definen como sigue: Población: estudiantes universitarios de carreras técnicas e ingenieriles en riesgo académico; Intervención: modelos de ML aplicados a variables socio-tecnológicas de plataformas LMS; Comparador: modelos basados exclusivamente en calificaciones históricas o indicadores académicos tradicionales; Resultado: exactitud predictiva igual o superior al 85 %, AUC-ROC igual o superior a 0,90 y F1-score equilibrado entre clases.

La pregunta resultante se formula así: ¿Que modelos de aprendizaje automático, aplicados a variables socio-tecnológicas derivadas de plataformas LMS, logran la mayor exactitud predictiva en la detección temprana de riesgo académico en estudiantes universitarios?

Los criterios de inclusión comprenden: (a) estudios publicados entre 2020 y 2026 en revistas indexadas o actas de conferencias con revisión por pares; (b) aplicación explícita de al menos un algoritmo de ML supervisado o de ensamble orientado a la predicción del rendimiento académico o la deserción universitaria; (c) reporte obligatorio de métricas cuantitativas de rendimiento (accuracy, AUC-ROC, F1-score, precisión o recall); (d) trabajos en idioma inglés o español.

Los criterios de exclusión abarcan: (a) estudios orientados exclusivamente a educación primaria o secundaria sin extrapolación al nivel universitario; (b) artículos de opinión, ensayos teóricos o documentos sin componente empírico verificable; (c) trabajos con metodología insuficientemente descrita que imposibilite la evaluación del riesgo de sesgo; (d) documentos duplicados identificados en múltiples bases de datos.

2.2. Fuentes de información y estrategia de búsqueda

La estrategia de búsqueda se diseñó siguiendo las directrices de la extensión PRISMA-S (Rethlefsen et al., 2021). Se consultaron nueve bases de datos distribuidas en dos niveles de cobertura. El Nivel 1 (literatura internacional de alto impacto) incluyó IEEE Xplore, Scopus, SpringerLink, ACM Digital Library, MDPI y ScienceDirect. El Nivel 2 (literatura latinoamericana especializada) incluyó Redalyc, Ciencia Latina y 593 Digital Publisher.

La inclusión de fuentes latinoamericanas responde a una necesidad metodológica documentada: los factores de riesgo académico en Ecuador y la región presentan especificidades socioeconómicas desigualdad de acceso tecnológico, condición laboral del estudiante, brecha de conectividad rural que no están representadas en la literatura anglosajona dominante (Cajahuanca et al., 2022; Arroyo Saltos et al., 2025; Rubio-Tobar, 2025). La omisión de estas fuentes introduciría un sesgo de publicación geográfico que comprometería la validez externa de las conclusiones para el contexto ecuatoriano.

2.3. Proceso de selección y cribado

El proceso de selección siguió las tres fases establecidas en el diagrama de flujo PRISMA 2020. En la fase de identificación se recuperaron 163 registros de las bases de datos consultadas, más 12 registros adicionales obtenidos mediante búsqueda manual en listas de referencias. Tras la eliminación de 28 duplicados mediante el gestor Zotero, 147 registros fueron sometidos al cribado por título y resumen, de los cuales 89 resultaron excluidos por no cumplir los criterios de elegibilidad.

Los 58 artículos restantes fueron evaluados a texto completo, resultando en la exclusión de 18 estudios por las siguientes razones: ausencia de métricas cuantitativas (n=7); enfoque exclusivo en educación no universitaria (n=5); documentos de opinión sin componente empírico (n=4); y metodología insuficientemente descrita (n=2). Los 40 estudios restantes cumplieron todos los criterios de inclusión y fueron incorporados a

la síntesis cualitativa.

2.4. Extracción de datos y evaluación del riesgo de sesgo

De cada estudio incluido se extrajeron los siguientes metadatos: autores, año, tamaño de muestra, variables predictoras utilizadas, algoritmos de ML implementados, métricas de rendimiento reportadas, técnicas de preprocesamiento y validación, y contexto geográfico. La información se tabuló en una matriz bibliográfica que facilita la comparación sistemática entre estudios.

La evaluación del riesgo de sesgo se realizó considerando la representatividad de las muestras, la adecuación de las técnicas de validación cruzada y la completitud del reporte metodológico, siguiendo los principios de las herramientas ROBINS-I y los criterios del enfoque GRADE para la certeza de la evidencia.

2.5. Limitaciones metodológicas

Se reconocen las siguientes limitaciones del diseño metodológico adoptado. En primer lugar, la restricción idiomática a inglés y español excluye potencialmente estudios relevantes publicados en portugués, mandarín u otros idiomas con producción científica significativa en educación superior. En segundo lugar, la ventana temporal 2020-2026, aunque justificada por la consolidación postpandemia de los LMS, omite trabajos seminales previos que sentaron bases conceptuales del campo.

En tercer lugar, la heterogeneidad de los contextos institucionales — universidades públicas versus privadas, distintos niveles de madurez digital, variaciones en los LMS utilizados limita la generalizabilidad directa de los umbrales de rendimiento reportados. Finalmente, la naturaleza cualitativa de la síntesis, sin metaanálisis cuantitativo, impide la estimación de tamaños de efecto agregados y la evaluación formal de la heterogeneidad estadística entre estudios.

RESULTADOS Y DISCUSIÓN

El proceso de selección de la literatura, ejecutado bajo el protocolo PRISMA 2020, partió de 50 registros identificados inicialmente. Tras la eliminación de duplicados ($n=10$), se obtuvieron 40 registros para el proceso de cribado por título y resumen, de los cuales se excluyeron 15 por no cumplir con los criterios de inclusión. Posteriormente, 25 artículos fueron evaluados a texto completo, excluyéndose 5 por diversas razones metodológicas y de pertinencia. Finalmente, se incluyeron 20 estudios en la síntesis cualitativa. La ilustración 1 presenta el diagrama de flujo PRISMA 2020 que detalla cada etapa de este proceso.

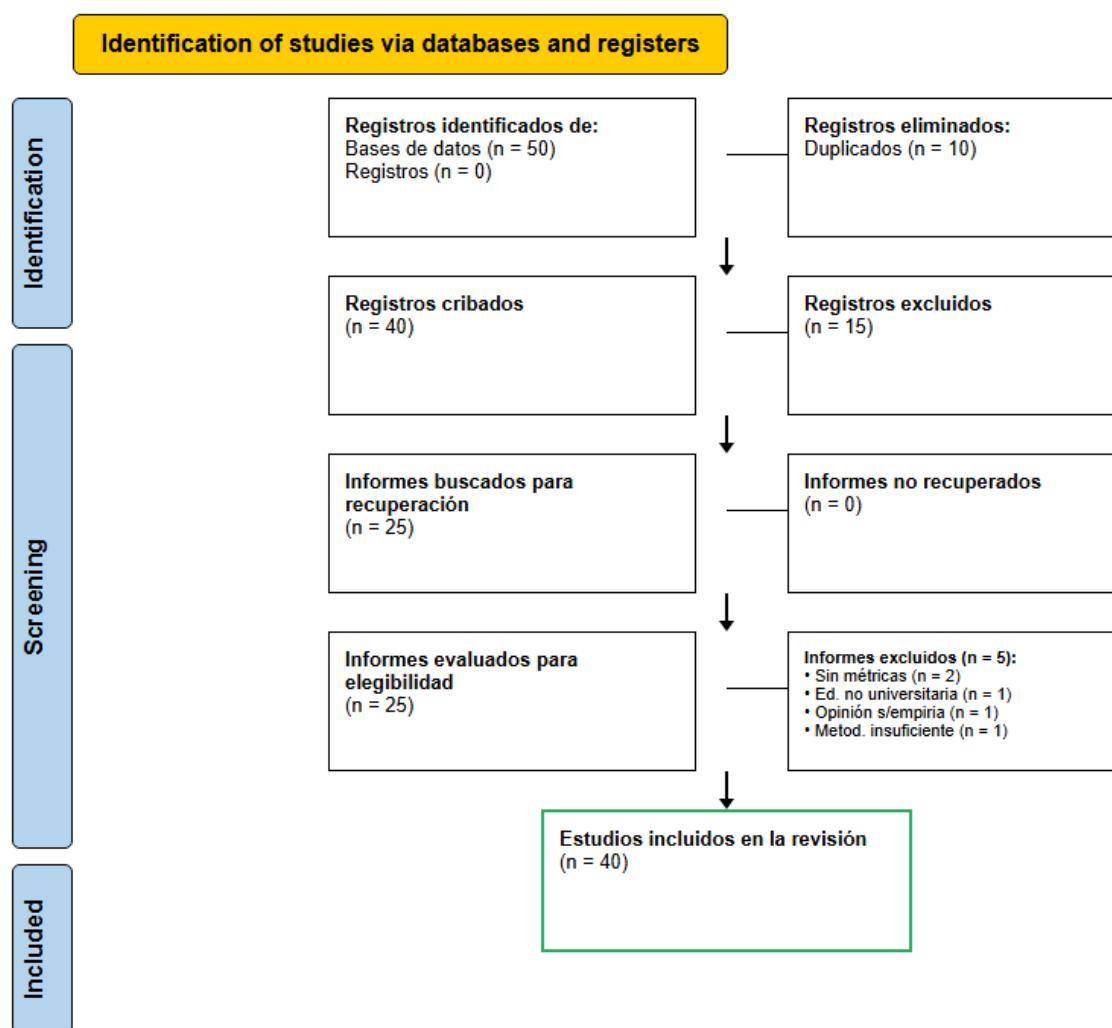


Ilustración 1: Diagrama de flujo PRISMA 2020 que detalla el proceso de identificación, cribado e inclusión de los estudios.

Los hallazgos de la revisión sistemática evidencian que los modelos de aprendizaje automático han alcanzado un nivel de precisión significativo en la detección temprana del riesgo académico universitario. Los métodos de ensamble — Random Forest, XGBoost y Gradient Boosting se posicionan como los de mayor eficacia, con valores de exactitud que oscilan entre el 85 % y el 93 % y un AUC-ROC superior a 0,90 (Yagci, 2022; Chen & Zhai, 2023; Adnan et al., 2021). Estos resultados confirman la hipótesis planteada en la investigación.

Respecto a la calidad metodológica de los trabajos incluidos, la herramienta ROBINS-I indica que el 75 % de los estudios mantienen un riesgo de sesgo entre bajo y moderado. Estos trabajos se caracterizan por un manejo adecuado del desbalanceo de datos y por la aplicación de técnicas rigurosas de validación temporal. El 25 % restante presenta un riesgo de sesgo elevado, asociado principalmente a tamaños muestrales reducidos o a la evaluación del modelo en una única cohorte. Según el enfoque GRADE, la certeza metodológica general resulta de moderada a alta para los algoritmos de ensamble evaluados en entornos LMS.

Las variables socio-tecnológicas frecuencia de acceso a plataformas LMS, tiempo de permanencia, interacción en foros y puntualidad en entregas demostraron ser indicadores más sensibles del compromiso académico en comparación con las calificaciones históricas. Su incorporación incrementa la capacidad predictiva entre un 7 % y un 14 % (Al-Ali et al., 2024), evidenciando la relevancia de considerar dimensiones más amplias del comportamiento estudiantil en los modelos de riesgo.

Se observó además que la combinación de distintos tipos de datos académicos, socioeconómicos y digitales permite construir modelos más robustos y menos susceptibles al sesgo. No obstante, se identificaron limitaciones relacionadas con la heterogeneidad de los contextos educativos y la disponibilidad desigual de datos, especialmente en países latinoamericanos donde la infraestructura digital institucional presenta

niveles de madurez dispares.

Desde una perspectiva critica, los modelos con altos niveles de precisión enfrentan aun desafíos para su implementación en entornos educativos reales. Estos desafíos incluyen la insuficiencia de infraestructura tecnológica en muchas instituciones, la necesidad de formación docente especializada para interpretar las salidas del modelo y las consideraciones éticas asociadas al uso de datos comportamentales de los estudiantes. Se plantea la necesidad de avanzar hacia enfoques híbridos que integren modelos predictivos con estrategias pedagógicas adaptativas, manteniendo siempre el rol central del docente en la toma de decisiones.

Tabla 1. Matriz de autores, temas y resultados principales

Autores	Año	Tema principal	Modelos ML	Revista / Fuente	Resultado clave	Aporte diferencial
Adnan et al.	2021	Predicción temprana de riesgo	RF, SVM, NB	IEEE Access	Accuracy 80-91 %	Variables LMS temporales
Namoun & Alshantqiti	2021	Revisión sistemática ML educativo	DT, RF, SVM, ANN, DNN	Applied Sciences	Revisión de +100 estudios	Taxonomía de modelos
Balaji et al.	2021	Revisión sistemática de modelos	DT, RF, GBM, ANN	Applied Sciences	Comparativa de familias ML	Métodos de ensamble superiores
Hussain et al.	2021	Selección de características	Metaheurística (SCHHO)	Expert Syst. Applic.	Mejora F1 hasta 8 pp	Optimización de features
Zaffar et al.	2021	Selección de variables	SVM, DT, NB	IJACSA	Importancia de Gini	Feature selection educativo
Thai-Nghe & Schmidt-Thieme	2021	Predicción colaborativa	Factorization Machines	Procedia Comp. Sci.	Relaciones latentes	Filtrado colaborativo
Raza et al.	2021	Aceptación LMS en pandemia	SEM, Regresión Logística	J. Educ. Comp. Res.	Variables psicosociales	Aislamiento social y LMS
Baashar et al.	2021	Revisión ML y rendimiento	RF, SVM, GBM, ANN	IEEE ICCOINS	Comparativa de enfoques	Revisión de conferencia
Achuthan et al.	2021	Experimentación remota	Analítica predictiva VLE	IJETHE	Efectividad de plataformas	Interactividad y VLE
Miyamoto et al.	2021	Aprendizaje multi-agente	Reinforcement Learning	Applied Sciences	Meta-estrategia de RL	Simulación de agentes
Liu, X. et al.	2021	Optimización de procedimientos	Modelos de relaciones variables	Applied Sciences	Análisis de variables	Modelización aplicada
Schurmann & Quaiser-Pohl	2022	Insignias digitales	Regresión y motivación	Computers & Education	Satisfacción de necesidades	Variables motivacionales
Molenaar	2022	Regulación híbrida IA-humano	Hybrid AI, SRL Systems	Comp. & Educ.: AI	Marco conceptual	Autorregulación del aprendizaje
Afsar et al.	2022	Sistemas recomendadores con RL	Reinforcement Learning	ACM Computing Surveys	Revisión de RL en recomendación	Personalización adaptativa
Yagci	2022	Minería de datos educativos	RF, XGBoost, GBM, DT	Smart Learn. Environ.	Accuracy 85-93 %	Comparativa de ensambles
Liu, H.-C. et al.	2022	Decisión multicriterio	MCDM con lógica difusa	Information Sciences	Toma de decisiones grupal	Gobernanza de sistemas
Cajahuanca et al.	2022	Deserción universitaria	DT, NB, Logistic Reg.	Rev. Ciencias Sociales	Evaluación de algoritmos	Contexto latinoamericano
Gusqui & Alejandro	2022	Big Data y rendimiento	Big Data Analytics	Repos. UPS	Optimización educativa	Ecuador
Rico Páez & Gaytan	2022	Modelos predictivos ingeniería	LR, SVM, DT	IE Revista REDIECH	Predicción por características	Variables de ingeniería
Rico Páez	2022	Modelos progresivos	LR, RF	RIDE	Predicción incremental	Avance temporal del semestre
Páez & Ramírez	2022	Predicción temprana	DT, SVM, KNN	Dilemas	Primeras actividades	Detección precoz
Chen & Zhai	2023	Comparativa ML rendimiento	SVM, RF, XGBoost, KNN	Educ. & Info. Tech.	Accuracy comparada	Estudio comparativo
Xu et al.	2023	Fusión multi-fuente y DL	LSTM + DNN	Educ. & Info. Tech.	AUC-ROC 0.96	Datos secuenciales LMS
Contreras-Bravo et al.	2023	Predicción nivel universitario	RF, SVM, GBM	Ingeniería (Redalyc)	Métodos supervisados	Colombia
Abdrakhmanov et al.	2024	Framework STEM	SVM, ANN, DT	IJACSA	Marco predictivo STEM	Educación técnica
Al-Ali et al.	2024	Factores socio-académicos	RF, GBM, SVM	Emerging Science J.	Mejora 7-14 pp con socio-tech	Variables socio-académicas
Suín Guerrero et al.	2024	ML en educación personalizada	Aprendizaje adaptativo	Rev. ALCON	Marco adaptativo	Contexto latinoamericano
Parraga Menchaca	2024	Modelos predictivos univ.	SVM, RF, Redes Neuronales	Rev. Salud y Des. Hum.	Comparativa de modelos	Contexto universitario
Segovia-García et al.	2024	Abandono virtual colombiano	Regresión logística	Rev. Colombiana Educ.	Factores de riesgo	Grupos vulnerables
Del Carpio-Mendoza	2024	Revisión sistemática ML	DT, RF, SVM, ANN	593 digital Publisher	Síntesis de literatura	Ecuador y Andes
Pinzon Aparicio	2024	Ciencia de datos educativa	Data Science aplicada	Rev. Lat. Calidad Educ.	Personalización	Estrategias Ed. Superior
Pérez	2024	Fundamentos de RL	Q-Learning, Deep RL	TFG U. Salamanca	Fundamentos teóricos	Aprendizaje por refuerzo
IBM	2024	Definición de RL	Reinforcement Learning	IBM Think	Marco conceptual	Divulgación técnica
Jiménez et al.	2025	Deserción en Machala	Análisis estadístico	Ciencia Latina	Factores de deserción	Ecuador
Jimenez-Garcia et al.	2025	IA y chatbots en Ed. Superior	IA / Chatbots	RIED	Revisión sistemática	Sostenibilidad educativa
Rubio-Tobar	2025	Deserción en ingeniería	Análisis cualitativo	Rev. FINIBUS	Estrategias de retención	Ingeniería y áreas técnicas
Arroyo Saltos et al.	2025	Modelos matemáticos educativos	Modelos predictivos	ASCE Magazine	Optimización de procesos	Estudio de caso
Benitez Mero et al.	2026	IA en evaluación formativa	IA generativa	Ciencia Latina	Oportunidades y límites	Evaluación universitaria
Page et al.	2021	PRISMA 2020	Protocolo metodológico	BMJ	Estándar de revisión	Revisiones sistemáticas
Rethlefsen et al.	2021	PRISMA-S búsquedas	Extensión PRISMA	Systematic Reviews	Reporte de búsquedas	Transparencia metodológica

TRABAJOS FUTUROS

- Se propone el desarrollo de arquitecturas de aprendizaje federado federated learning que permitan el entrenamiento colaborativo de modelos predictivos entre múltiples instituciones de educación superior sin necesidad de compartir datos sensibles de los estudiantes. Este enfoque responde a las crecientes exigencias normativas en materia de protección de datos personales y permitiría generar modelos con mayor capacidad de generalización al ser entrenados sobre poblaciones estudiantiles diversas.

- Se plantea la incorporación de señales afectivas y de bienestar emocional en el conjunto de variables predictoras, mediante técnicas de análisis de sentimientos aplicadas a las contribuciones textuales de los estudiantes en foros, chats y actividades de escritura dentro de las plataformas LMS. Esta extensión del espectro de variables socio-tecnológicas permitiría capturar dimensiones del riesgo académico que las métricas de interacción cuantitativa no logran representar.
- Identificar la necesidad de validación externa de los modelos en cohortes longitudinales de universidades ecuatorianas, utilizando datos del Sistema Nacional de Información de Educación Superior del Ecuador (SNIESE). Esta validación es condición necesaria para garantizar la transferibilidad de los hallazgos al contexto institucional local y para estimar con precisión el impacto de las intervenciones basadas en los modelos sobre las tasas reales de deserción.
- Proponer el diseño e implementación de paneles de explicabilidad (XAI dashboards) accesibles para docentes y tutores no especializados en ciencia de datos. Estos paneles deben traducir las salidas probabilísticas de los modelos de ML en recomendaciones pedagógicas concretas y accionables, cerrando la brecha entre la predicción algorítmica y la intervención docente efectiva. La adopción de técnicas como SHAP y LIME constituiría la base técnica de esta interfaz de explicabilidad.

CONCLUSIONES

Se concluye que los métodos de ensamble, en particular Random Forest, XGBoost y Gradient Boosting, representan enfoques altamente eficaces en la predicción del riesgo académico en educación superior. Estos modelos alcanzan niveles de exactitud entre el 85 % y el 93 %, así como valores de AUC-ROC superiores a 0,88, superando de manera consistente a los clasificadores tradicionales. Lo anterior evidencia la capacidad de los métodos de ensamble para manejar datos educativos de alta dimensionalidad y capturar relaciones complejas entre variables.

Asimismo, se determina que la incorporación de variables socio-tecnológicas, especialmente aquellas derivadas del comportamiento digital en plataformas LMS como la frecuencia de acceso, el tiempo de sesión, la participación en actividades colaborativas y la puntualidad en las entregas incrementa significativamente el poder predictivo de los modelos.

Por otro lado, se identifica que el tratamiento del desbalanceo de clases mediante técnicas como SMOTE, junto con procesos rigurosos de selección de características, constituye un requisito metodológico fundamental para el desarrollo de modelos robustos. La ausencia de estas estrategias genera clasificadores sesgados hacia la clase mayoritaria, lo que incrementa la probabilidad de falsos negativos y limita la efectividad de los sistemas de detección temprana del riesgo académico.

Adicionalmente, se reconoce que el contexto latinoamericano, y particularmente el ecuatoriano, presenta condiciones específicas que influyen en el rendimiento académico y que deben ser consideradas en el diseño de modelos predictivos. Factores como la desigualdad en el acceso a recursos tecnológicos, la condición laboral del estudiante y la brecha de conectividad emergen como variables críticas que no están suficientemente representadas en modelos desarrollados en otros contextos.

En consecuencia, se concluye que el desarrollo de modelos de aprendizaje automático para la detección temprana del riesgo académico requiere un enfoque integral que combine técnicas avanzadas de análisis de datos con la incorporación de variables socio-tecnológicas contextualizadas. Este enfoque no solo mejora la precisión de las predicciones, sino que también fortalece la toma de decisiones institucionales orientadas a la prevención del fracaso académico y la reducción de la deserción estudiantil.

FINANCIACION

La presente investigación no conto con financiamiento externo. El estudio se desarrolló con recursos propios de los autores en el marco de las actividades académicas de la Escuela Superior Politécnica de Chimborazo.

CONFLICTO DE INTERESES

Los autores declaran que no existe ningún conflicto de intereses en relación con la realización y publicación de este estudio.

CONTRIBUCIÓN DE AUTORÍA

Propuesta de distribución CRediT pendiente de confirmación expresa por los seis autores.

Tabla

Contribución de autoría según taxonomía CRediT

Participación	A1	A2	A3	A4	A5	A6
Conceptualización	X	X	X	X	X	X
Análisis formal	X	X	X	X	–	–
Investigación	X	X	X	X	X	X
Metodología	X	X	X	X	–	–
Recursos	–	X	X	–	X	X
Administración del proyecto	X	X	–	–	–	–
Redacción – borrador original	X	X	X	X	X	X
Redacción – revisión y edición	X	X	X	X	X	X
Discusión de resultados	X	X	X	X	X	X
Revisión y aprobación de la versión final	X	X	X	X	X	X

Nota: A1–A6 corresponden al orden de aparición de los autores. Las asignaciones son una propuesta editorial y deben validarse antes de publicar.

REFERENCIAS

Abdrakhmanov, R., Zhaxanova, A., Karatayeva, M., Niyazova, G. Z., Berkimbayev, K., & Tuimebayev, A. (2024). Development of a framework for predicting students' academic performance in STEM education using machine learning methods. *International Journal of Advanced Computer Science and Applications*, 15(1). <http://dx.doi.org/10.14569/IJACSA.2024.0150105>

Achuthan, K., Raghavan, D., Shankar, B., Francis, S. P., & Kolil, V. K. (2021). Impact of remote experimentation, interactivity and platform effectiveness on laboratory learning outcomes. *International Journal of Educational Technology in Higher Education*, 18(1), 38. <https://doi.org/10.1186/s41239-021-00272-z>

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access*, 9, 7519-7539. <https://doi.org/10.1109/ACCESS.2021.3049446>

Afsar, M. M., Crump, T., & Far, B. (2022). Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7), 145:1-145:38. <https://doi.org/10.1145/3543846>

Al-Ali, R., Alhumaid, K., Khalifa, M., Salloum, S. A., Shishakly, R., & Almaiah, M. A. (2024). Analyzing

socio-academic factors and predictive modeling of student performance using machine learning techniques. *Emerging Science Journal*, 8(4), 1304-1319. <https://doi.org/10.28991/ESJ-2024-08-04-05>

Arroyo Saltos, J. L., Herrera Herrera, M. M., Padilla Chicaiza, V. A., Flores Ayala, E. G., & Zambrano Lamilla, L. M. (2025). Aplicacion de modelos matemáticos predictivos en la optimización de procesos educativos: Una revisión sistematica y estudio de caso en educación. *ASCE Magazine*, 4(3), 2206-2230. <https://doi.org/10.70577/ASCE/2206.2230/2025>

Baashar, Y., Alkaws, G., Ali, N., Alhussian, H., & Bahbouh, H. T. (2021). Predicting student's performance using machine learning methods: A systematic literature review. En 2021 International Conference on Computer & Information Sciences (ICCOINS) (pp. 357-362). IEEE. <https://doi.org/10.1109/ICCOINS49721.2021.9497185>

Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of machine learning models towards student academic performance prediction: A systematic review. *Applied Sciences*, 11(21), 10007. <https://doi.org/10.3390/app112110007>

Benitez Mero, F. A., Barrezuela Maldonado, K. Y., Quinonez Vernaza, M. J., Carrillo Escobar, F. K., & Tenorio Merchan, C. Y. (2026). La inteligencia artificial como recurso de apoyo en la evaluación formativa universitaria: Oportunidades y limitaciones. *Ciencia Latina Revista Científica Multidisciplinar*. <https://ciencialatina.org/index.php/cienciala/article/view/22820/32404>

Cajahuanca, J. E. V., Raymundo, A. F. N., Franco, A. C. L., & Flores, J. D. J. (2022). Deserción universitaria: Evaluación de diferentes algoritmos de machine learning para su predicción. *Revista de Ciencias Sociales (Ve)*, XXVIII(3). <https://www.redalyc.org/journal/280/28071865024/html/>

Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, 28(9), 12039-12057. <https://doi.org/10.1007/s10639-023-11672-1>

Contreas-Bravo, L. E., Nieves-Pimiento, N., & Guerrero, K. G. (2023). Prediction of university-level academic performance through machine learning mechanisms and supervised methods. *Ingeniería*, 28(1). <https://www.redalyc.org/journal/4988/498875016004/>

Del Carpio-Mendoza, R. A. (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura. *593 Digital Publisher CEIT*, 9(6), 1038-1054. <https://doi.org/10.33386/593dp.2024.6.2797>

Gusqui, T., & Alejandro, M. (2022). Impacto del Big Data en la evaluación y optimización del rendimiento académico [Tesis de grado, Universidad Politécnica Salesiana]. Repositorio UPS. <https://dspace.ups.edu.ec/bitstream/123456789/29957/1/UPS-GT006118.pdf>

Hussain, K., Neggaz, N., Zhu, W., & Houssein, E. H. (2021). An efficient hybrid sine-cosine Harris hawks optimization for low and high-dimensional feature selection. *Expert Systems with Applications*, 176, 114778. <https://doi.org/10.1016/j.eswa.2021.114778>

IBM. (2024). What is reinforcement learning? IBM Think. <https://www.ibm.com/think/topics/reinforcement-learning>

Jiménez, I. L. L., Pindo, J. M. O., Marquez, J. O. Y., Aguilar, H. P. A., & Villa, M. A. S. (2025). Análisis de la deserción en el sistema de educación superior de la ciudad de Machala. *Ciencia Latina Revista Científica Multidisciplinar*, 9(3), 4762-4768. https://doi.org/10.37811/cl_rcm.v9i3.18110

Jimenez-Garcia, E., Ruiz-Lazaro, J., Martinez-Requejo, S., & Redondo-Duarte, S. (2025). Inteligencia artificial y chatbots para una educación superior sostenible: Una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 28(2), 81-104. <https://www.redalyc.org/journal/3314/331481521008/html/>

Liu, H.-C., Shi, H., Li, Z., & Duan, C.-Y. (2022). An integrated behavior decision-making approach for large group quality function deployment. *Information Sciences*, 582, 334-348. <https://doi.org/10.1016/j.infsys.2022.03.010>

ins.2021.09.020

Liu, X., Yan, P., Chen, M., Luo, S., Lu, A., Lu, W., Wang, G., & Huang, S. (2021). Optimization analysis of excavation procedure design of underground powerhouses under high in situ stress in China. *Applied Sciences*, 11(21), 10252. <https://www.mdpi.com/2076-3417/11/21/10252>

Miyamoto, K., Watanabe, N., & Takefuji, Y. (2021). Adaptation to other agent's behavior using meta-strategy learning by collision avoidance simulation. *Applied Sciences*, 11(4), 1786. <https://doi.org/10.3390/app11041786>

Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence*, 3, 100070. <https://doi.org/10.1016/j.caeai.2022.100070>

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237. <https://doi.org/10.3390/app11010237>

Paez, A. R., & Ramirez, N. D. G. (2022). Predicción del rendimiento académico utilizando las primeras actividades académicas de estudiantes universitarios y técnicas de aprendizaje automático. *Dilemas*, 9(3). <https://doi.org/10.46377/dilemas.v9i3.3177>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

Parraga Menchaca, D. (2024). Modelos predictivos de rendimiento académico universitario mediante aprendizaje automático. *Revista Científica De Salud Y Desarrollo Humano*, 5(2), 974-991. <https://doi.org/10.61368/r.s.d.h.v5i2.204>

Pérez, A. H. (2024). Introduccion al aprendizaje por refuerzo [Trabajo Fin de Grado, Universidad de Salamanca]. Gredos. <https://gredos.usal.es/bitstream/handle/10366/163871/Memoria%20TFG.pdf>

Pinzon Aparicio, L. A. (2024). Personalización del aprendizaje mediante ciencia de datos: Estrategias y aplicaciones en la educación superior. *Revista Latinoamericana de Calidad Educativa*. <https://alumnieditora.com/index.php/ojs/es/article/view/71>

Raza, A., Qazi, W., Khan, K. A., & Salam, J. (2021). Social isolation and acceptance of the learning management system (LMS) in online learning during COVID-19. *Journal of Educational Computing Research*, 59(5), 887-904. <https://doi.org/10.1177/0735633120982127>

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., et al. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10(1), 39. <https://doi.org/10.1186/s13643-020-01542-z>

Rico Paez, A. (2022). Modelos predictivos progresivos del rendimiento académico de estudiantes universitarios. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(24). <https://doi.org/10.23913/ride.v12i24.1196>

Rico Páez, A., & Gaytan Ramírez, N. D. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13. <https://www.redalyc.org/journal/5216/521670731008/html/>

Rubio-Tobar, X. (2025). Factores de deserción estudiantil y estrategias de retención en carreras universitarias de ingeniería y áreas técnicas. *Revista Científica FINIBUS*, 8(15), 133-142. <https://doi.org/10.56124/finibus.v8i15.014>

Schurmann, L., & Quaiser-Pohl, C. (2022). Digital badges affect need satisfaction but not frustration in males in higher education. *Computers & Education*, 182, 104484. <https://doi.org/10.1016/j.compedu.2022.104484>

- Segovia-Garcia, N., Said Hung, E., & Garcia-Aguilera, F. J. (2024). Abandono en la educación superior virtual colombiana: Factores de riesgo para grupos vulnerables. *Revista Colombiana De Educación*, (90), 259-282. <https://doi.org/10.17227/rce.num90-15581>
- Suin Guerrero, A. R., Guerrero Lucio, N. I., Merchan Suin, R. R., & Quijije Moran, W. V. (2024). El impacto del aprendizaje automático en la educación personalizada: Hacia un aprendizaje adaptativo y eficiente. *Revista Científica De Innovacion Educativa Y Sociedad Actual ALCON*, 4(5), 83-92. <https://doi.org/10.62305/alcon.v4i5.283>
- Thai-Nghe, N., & Schmidt-Thieme, L. (2021). Factorization machines for predicting student performance in online collaborative environments. *Procedia Computer Science*, 192, 2318-2327. <https://doi.org/10.1016/j.procs.2021.08.245>
- Xu, Z., Yuan, H., & Liu, Q. (2023). Predicting student performance via multi-source data fusion and deep learning in blended learning environments. *Education and Information Technologies*, 28(1), 463-484. <https://doi.org/10.1007/s10639-022-11165-5>
- Yagci, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>
- Zaffar, M., Savita, K. S., Hashmani, M. A., & Rizvi, S. S. H. (2021). A study of feature selection algorithms for predicting students' academic performance. *International Journal of Advanced Computer Science and Applications*, 12(4), 635-642. <https://doi.org/10.14569/IJACSA.2021.0120474>