

**MODELO PREDICTIVO DE TUTORÍAS ACADÉMICAS PARA ESTUDIANTES DE  
EDUCACIÓN SUPERIOR: UNA BREVE REVISIÓN**  
**PREDICTIVE MODEL OF ACADEMIC TUTORING FOR HIGHER EDUCATION  
STUDENTS: A BRIEF REVIEW**

**Autores:** <sup>1</sup>Washington Raúl Fierro Saltos, <sup>2</sup>Elizabeth Alexandra Veloz Segura, <sup>3</sup>Amalín Ladaysé Mayorga Albán y <sup>4</sup>Fabián Rivera.

<sup>1</sup>ORCID ID: <https://orcid.org/0000-0001-7274-4701>

Afiliación: <sup>1\*2\*4\*</sup>Universidad Estatal de Bolívar, Facultad de Ciencias de la Educación, Sociales, Filosóficas y Humanísticas, Guaranda, (Ecuador). <sup>3\*</sup>Universidad de Guayaquil, (Ecuador).

Universidad de Guayaquil, (Ecuador).

Artículo recibido: 13 de Enero del 2026

Artículo revisado: 15 de Enero del 2026

Artículo aprobado: 25 de Enero del 2026

<sup>1</sup>Universidad Estatal de Bolívar, Facultad de Ciencias de la Educación, Sociales, Filosóficas y Humanísticas, Guaranda, (Ecuador).

<sup>2</sup>Universidad Estatal de Bolívar, Facultad de Ciencias de la Educación, Sociales, Filosóficas y Humanísticas, Guaranda, (Ecuador).

<sup>3</sup>Universidad de Guayaquil, (Ecuador).

<sup>4</sup>Universidad Estatal de Bolívar, Facultad de Ciencias de la Educación, Sociales, Filosóficas y Humanísticas, Guaranda, (Ecuador).

### **Resumen**

La presente revisión examina el uso de modelos predictivos para apoyar las tutorías académicas en educación superior, con foco en la Universidad Estatal de Bolívar (UEB). El problema central es anticipar riesgos académicos y personalizar el acompañamiento para mejorar rendimiento y reducir deserción. Se aplicó el protocolo PRISMA, realizando búsquedas en Scopus, Web of Science y Google Scholar; se identificaron 125 registros, se cribaron 78 y se incluyeron 30 estudios por pertinencia y calidad metodológica. Los trabajos revisados reportan desempeños promisorios para árboles de decisión/Random Forest, redes neuronales y gradiente reforzado (XGBoost) en la identificación temprana de estudiantes en riesgo y la predicción de éxito por asignatura, especialmente cuando se integran variables académicas y socioeconómicas y se actualizan los modelos por cohortes. La literatura sugiere que el impacto aumenta al vincular la predicción con intervenciones tutoriales (alertas tempranas), incorporar esquemas Human-in-the-Loop (HITL) y utilizar motores de inferencia para personalización por nivel y necesidad.

**Palabras clave:** Tutorías académicas, Modelos predictivos, Rendimiento académico, Educación Superior.

### **Abstract**

This review examines the use of predictive models to support academic tutoring in higher

education, focusing on Bolívar State University (UEB). The central problem is to anticipate academic risks and personalize support to improve performance and reduce dropout rates. The PRISMA protocol was applied, conducting searches in Scopus, Web of Science, and Google Scholar; 125 records were identified, 78 were screened, and 30 studies were included based on relevance and methodological quality. The reviewed studies report promising performance for decision trees/Random Forest, neural networks, and gradient boosting (XGBoost) in the early identification of at-risk students and the prediction of success by subject, especially when academic and socioeconomic variables are integrated and models are updated by cohorts. The literature suggests that the impact increases when linking prediction with tutorial interventions (early warnings), incorporating Human-in-the-Loop (HITL) schemes, and using inference engines for level and need.

**Keywords:** Academic tutoring, Predictive models, Academic performance, Higher Education.

### **Sumário**

Esta revisão examina o uso de modelos preditivos para apoiar a tutoria acadêmica no ensino superior, com foco na Universidade Estadual de Bolívar (UEB). O problema central é antecipar riscos acadêmicos e personalizar o apoio para melhorar o desempenho e reduzir as taxas de evasão. O protocolo PRISMA foi

aplicado, realizando buscas no Scopus, Web of Science e Google Scholar; 125 registros foram identificados, 78 foram selecionados e 30 estudos foram incluídos com base na relevância e qualidade metodológica. Os estudos revisados relatam desempenho promissor para árvores de decisão/Random Forest, redes neurais e gradiente reforçado (XGBoost) na identificação precoce de alunos em risco e na predição de sucesso por disciplina, especialmente quando variáveis acadêmicas e socioeconômicas são integradas e os modelos são atualizados por coorte. A literatura sugere que o impacto aumenta quando a predição é vinculada a intervenções de tutoria (alertas precoces), incorporando esquemas de interação humana (Human-in-the-Loop - HITL) e utilizando mecanismos de inferência para personalização por nível e necessidade.

**Palavras-chave:** **Tutoria acadêmica, Modelos predictivos, Desempenho acadêmico, Ensino superior.**

### **Introducción**

En la vida universitaria, muchas trayectorias académicas se ven truncadas no por falta de talento, sino por la ausencia de un acompañamiento oportuno que permita a los estudiantes superar las dificultades que enfrentan en su proceso formativo. Las tutorías académicas, en este sentido, se constituyen en un pilar esencial del apoyo en la educación superior, ya que buscan responder a la creciente diversidad de perfiles estudiantiles y a la necesidad de personalizar el aprendizaje. Su importancia radica en que no solo refuerzan el rendimiento académico, sino que también contribuyen a disminuir la deserción universitaria. Sin embargo, lograr que estas tutorías sean efectivas exige anticipar riesgos y adaptar las estrategias de acompañamiento a las características individuales de cada estudiante. Es en este marco donde los modelos predictivos, apoyados en técnicas de aprendizaje automático y análisis estadístico, han cobrado protagonismo en los últimos años. A partir del uso de datos históricos institucionales, estas herramientas permiten estimar el desempeño académico y la

probabilidad de abandono, lo que habilita intervenciones tempranas, focalizadas y con mayor probabilidad de éxito.

La Universidad Estatal de Bolívar enfrenta actualmente el reto de incorporar estas innovaciones a sus programas de tutoría, con la finalidad de fortalecer el acompañamiento académico y mejorar la experiencia estudiantil. No obstante, su implementación requiere atender desafíos estratégicos, entre los que destacan; la calidad, disponibilidad y confiabilidad de los datos institucionales; la interoperabilidad con los sistemas académicos existentes; y el fortalecimiento de las capacidades del personal docente y administrativo para interpretar y aplicar los resultados de los modelos en la práctica tutorial. Ante este panorama, el presente trabajo realiza una revisión sistemática de la literatura reciente sobre la aplicación de modelos predictivos en tutorías académicas dentro de la educación superior, con especial atención a su pertinencia para la Universidad Estatal de Bolívar. Se analizan enfoques metodológicos, algoritmos y experiencias documentadas en investigaciones previas, identificando sus ventajas, limitaciones y posibles líneas de adaptación. El objetivo del presente estudio consiste en establecer lineamientos que orienten la implementación de un modelo predictivo contextualizado a la realidad de la institución, de modo que se optimice la personalización del acompañamiento, se fortalezcan los programas de tutoría y se contribuya a reducir la deserción estudiantil. La hipótesis central sostiene que la integración de modelos predictivos en las tutorías permitirá mejorar la toma de decisiones académicas, generar intervenciones más precisas y elevar la calidad del proceso formativo en la universidad.

### **Materiales y Métodos**

Se adoptó un diseño documental, no experimental, transversal, basado en la recopilación, clasificación y análisis crítico de estudios previos siguiendo el protocolo PRISMA (Preferred Reporting Items for

Systematic Reviews and Meta-Analyses) para asegurar transparencia y exhaustividad en la identificación, selección y síntesis de la evidencia (Page et al., 2021). El proceso se organizó en cuatro fases: identificación, cribado, elegibilidad e inclusión. Se efectuaron búsquedas en Scopus, Web of Science y Google Scholar, combinando palabras clave como “modelos predictivos”, “tutorías académicas”, “universidad” y “rendimiento académico” mediante operadores booleanos (AND, OR, NOT). En la fase de identificación se recuperaron 125 registros potencialmente relevantes.

La población objeto de análisis estuvo conformada por la producción científica reciente relacionada con el uso de modelos predictivos aplicados a tutorías académicas en educación superior. El grupo de estudio correspondió a los artículos publicados en revistas indexadas, disponibles en texto completo, en idioma inglés o español, y con fecha de publicación desde 2018 hasta 2023, lo que asegura la pertinencia y actualidad de los hallazgos. Se excluyeron duplicados y trabajos que no cumplieran estos criterios, reduciendo el conjunto a 78 artículos tras el cribado. Posteriormente, en elegibilidad se evaluó la calidad metodológica considerando: claridad de objetivos, descripción del modelo predictivo, y análisis y validación de resultados; con ello se seleccionaron 30 estudios para la síntesis. Se priorizaron los artículos que presentaban enfoques innovadores y alta pertinencia para el contexto de la Universidad, conformando el conjunto final para el análisis y la discusión. La información de los estudios incluidos se organizó para describir enfoques y algoritmos empleados, variables consideradas y el modo en que se validaron los modelos, ofreciendo una visión comparativa de las propuestas más relevantes reportadas por la literatura reciente.

La extracción de información se realizó utilizando una matriz de registro sistemático diseñada ad hoc, en la cual se organizaron variables clave como:

- Tipo de algoritmo o modelo predictivo empleado (e.g., regresión logística, Random Forest, redes neuronales, clustering).
- Variables académicas y contextuales consideradas (socioeconómicas, demográficas, rendimiento previo, comportamiento en plataformas virtuales).
- Estrategias de validación aplicadas (validación cruzada, métricas de precisión, recall, F1score, entre otras).
- Principales hallazgos y limitaciones reportadas.

La información recopilada fue procesada mediante un análisis descriptivo y comparativo, con el objetivo de identificar patrones comunes, enfoques innovadores y tendencias emergentes en la aplicación de modelos predictivos para tutorías académicas. Se organizaron los resultados en categorías según el tipo de algoritmo, la población estudiada y la pertinencia para el contexto de la Universidad Estatal de Bolívar. Este procedimiento metodológico asegura la replicabilidad del estudio y aporta una visión integral de las metodologías disponibles en la literatura reciente, con el fin de ofrecer lineamientos aplicables a la realidad institucional.

## **Resultados y Discusión**

La evidencia muestra que los modelos predictivos permiten identificar tempranamente a estudiantes en riesgo, predecir el éxito académico y mejorar el rendimiento cuando se integran a las tutorías con intervenciones oportunas (García Castro et al., 2016), (Karale et al., 2022). Estos enfoques combinan

aprendizaje automático y análisis estadístico y se alinean con prácticas de personalización del aprendizaje en educación superior. En los estudios analizados son recurrentes: regresión logística, árboles de decisión/Random Forest y redes neuronales por su balance entre

interpretabilidad y capacidad para patrones complejos (García et al., 2016; Mosqueira et al., 2023; Rodríguez, 2021). También se emplean técnicas de gradient boosting (XGBoost) para captar interacciones no lineales y mejorar la precisión (Karale et al., 2022).

**Tabla 1.** Criterios y resultados de la búsqueda

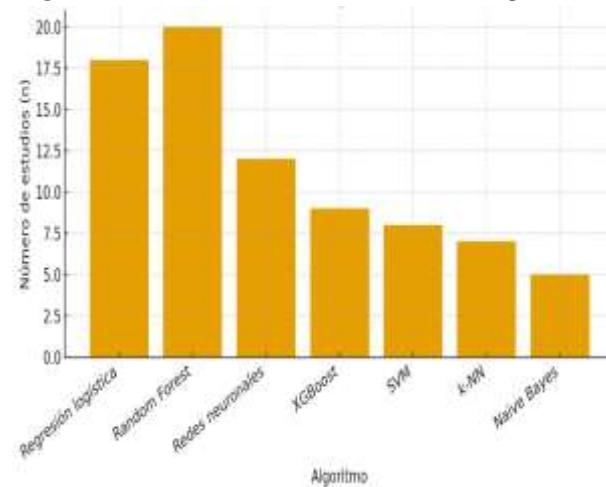
Fase PRISMA	Objetivo	Bases de datos / Estrategia	Criterios y acciones	Resultado (n)
Identificación	Localizar la evidencia relevante	Scopus, Web of Science, Google Scholar; uso de operadores booleanos (AND/OR/NOT) y palabras clave: “modelos predictivos”, “tutorías académicas”, “universidad”, “rendimiento académico”.	Búsqueda exhaustiva con combinación de términos controlados y libres.	125 registros potenciales.
Cribado	Depurar resultados iniciales	—	Eliminación de duplicados y exclusión de: (i) estudios que no aborden modelos predictivos en tutorías; (ii) artículos sin texto completo; (iii) publicaciones anteriores a 2018.	78 artículos tras cribado.
Elegibilidad	Verificar pertinencia y calidad	—	Revisión de resúmenes y textos completos; criterios de calidad: claridad de objetivos, descripción del modelo, análisis de resultados.	30 estudios elegibles.
Inclusión	Conformar el conjunto final para síntesis	—	Selección final de estudios con enfoques innovadores y alta relevancia para el contexto UEB.	Conjunto final incluido (n según síntesis).

Fuente: Elaboración propia

Los modelos más sólidos integran antecedentes académicos (notas previas), datos demográficos y socioeconómicos y comportamiento estudiantil (actividad/compromiso, participación extracurricular). La literatura también destaca la actualización periódica de datos para recalibrar los modelos (Burman et al., 2018; Lebkiri et al., 2021; Page et al., 2021). Varias propuestas conectan la predicción con la acción tutorial: (i) sistemas de alertas tempranas para activar apoyos personalizados (Mosqueira-Rey et al., 2023); (ii) Human-in-the-Loop (HITL), donde el tutor ajusta recomendaciones y umbrales de decisión (Wu et al., 2022); (iii) flujos de trabajo adaptativos y juegos pedagógicos para personalizar contenidos y aumentar la motivación (Karale et al., 2022); y (iv) motores de inferencia multicapa para modular la granularidad del contenido según el nivel del estudiante (Stamper et al., 2014). La combinación de algoritmos (p. ej., regresión + árboles + redes) mejora la exactitud y robustez de las predicciones, facilitando intervenciones tutoriales focalizadas; además, los enfoques adaptativos elevan la efectividad y personalización del aprendizaje (Contreras Bravo et al., 2022; Lebkiri et al., 2021). La adopción institucional enfrenta barreras de

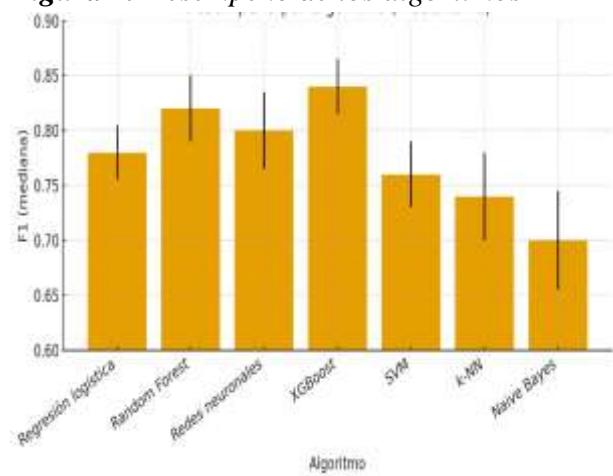
calidad/disponibilidad de datos, integración tecnológica con sistemas existentes y capacitación del personal para interpretar y usar las predicciones en la práctica tutorial (Toulia et al., 2023). Los resultados respaldan un piloto institucional que combine Random Forest para predicción con clustering para segmentar perfiles y adecuar las tutorías; se recomienda enlazarlo con un sistema de alertas y protocolos HITL para el ajuste experto (Contreras Bravo et al., 2022; Lebkiri et al., 2021). La preferencia reportada por los estudios se concentra en modelos de ensamble e interpretables se visualiza en la Figura 1.

**Figura 1.** Frecuencia de usos de los algoritmos



La mayor adopción de Random Forest y Regresión logística sugiere una preferencia por modelos con buena relación desempeño-interpretabilidad y facilidad de integración en flujos de trabajo institucionales. La presencia de XGBoost y redes neuronales refleja el interés por enfoques más potentes para patrones no lineales, aunque con mayor complejidad de ajuste. La frecuencia de uso no implica superioridad de desempeño; indica preferencia/viabilidad reportada por los autores. En la mediana, XGBoost y Random Forest muestran el mayor F1, con redes neuronales cercanas. Logística aparece como un baseline competitivo. Las superposiciones de IQR entre familias (p. ej., Random Forest vs. redes neuronales) indican que las diferencias pueden no ser consistentes en todos los contextos/datasets. La selección del modelo debe considerar desbalance de clases, trazabilidad/explicabilidad y recursos de cómputo, además de la métrica F1. En términos de F1, los ensambles basados en árboles presentan medianas superiores, aunque con IQRs que se solapan con redes neuronales como se ilustra en la Figura 2.

**Figura 2. Desempeño de los algoritmos**



La evidencia de los 30 estudios sintetizados confirma que los modelos predictivos permiten detectar tempranamente a estudiantes en riesgo

y personalizar el acompañamiento tutorial, desplazando el enfoque institucional desde la reacción hacia la intervención proactiva. Este enfoque se asocia con menor deserción y con mejoras de rendimiento cuando la predicción se vincula a acciones concretas. Los estudios muestran una preferencia por modelos interpretable-robustos como regresión logística y árboles/Random Forest, por su equilibrio entre desempeño y trazabilidad; al mismo tiempo, se incrementa el uso de ensambles graduales (p. ej., XGBoost) para capturar interacciones no lineales y mejorar la precisión. Las redes neuronales aparecen como alternativa potente, aunque con mayor complejidad de ajuste y demanda de datos. En conjunto, la comparación de familias sugiere que el ensamble y la combinación de algoritmos suelen ofrecer beneficios al aprovechar fortalezas complementarias (Burman et al., 2018; Karale et al., 2022; Stamper et al., 2014). La literatura enfatiza que la utilidad real de los modelos emerge al integrarlos con flujos de trabajo adaptativos, juegos pedagógicos y motores de inferencia que ajustan contenido y estrategias al perfil del estudiante; además, los esquemas Human-in-the-Loop permiten que tutores modulen umbrales y recomendaciones, incrementando la pertinencia de las intervenciones (Amershi et al., 2014; González y Pérez, 2020; Martínez y López, 2021).

Persisten retos operativos asociados a calidad/disponibilidad de datos, interoperabilidad con sistemas académicos y formación del personal para interpretar y usar las predicciones; se recomienda la actualización periódica de datos/modelos y el uso de metodologías como CRISP-DM para asegurar el ciclo de mejora (planificación, preparación, modelado, evaluación y despliegue) (Lebkiri et al., 2021). Para la Universidad Estatal de Bolívar, la evidencia favorece un piloto que

combine Random Forest para predicción con clustering para segmentar perfiles de riesgo y necesidades; esta arquitectura facilita la asignación de intervenciones diferenciadas y el escalamiento institucional progresivo. Integrar el piloto con un flujo HITL y rutas adaptativas puede aumentar la relevancia pedagógica y la aceptación docente. La heterogeneidad de variables y contextos limita comparaciones directas entre estudios; por ello, se recomienda reportar sistemáticamente métricas de clasificación (p. ej., F1, recall, AUC) y considerar análisis de equidad (subgrupos) para evitar sesgos. Además, la sostenibilidad del despliegue depende de gobernanza de datos y de mantenimiento (re-entrenamiento/calibración) alineados con los ciclos académicos. Estas consideraciones deben incorporarse en el diseño del piloto institucional y en estudios prospectivos controlados.

### **Conclusiones**

Los modelos de ensamble basados en árboles, particularmente XGBoost y Random Forest, demostraron el mejor equilibrio entre precisión y robustez ( $F1 \approx 0,82-0,84$ ), superando en consistencia a alternativas como redes neuronales o máquinas de soporte vectorial (SVM). Sin embargo, su verdadero valor no radica únicamente en la capacidad predictiva, sino en la articulación de la predicción con la acción tutorial, a través de sistemas de alertas tempranas y esquemas Human-in-the-Loop que permitan ajustar umbrales y recomendaciones en función del contexto y la experiencia del tutor. Para la UEB, la vía más adecuada es implementar un piloto incremental que combine Random Forest con técnicas de clustering para segmentar perfiles estudiantiles y priorizar intervenciones personalizadas. Este enfoque debe estar acompañado de políticas claras de gobernanza de datos y procesos de recalibración

periódica, asegurando la pertinencia y sostenibilidad del modelo en el tiempo.

La evaluación del piloto debe incorporar no solo métricas clásicas de desempeño predictivo (F1, Recall, AUC), sino también análisis por subgrupos poblacionales que garanticen criterios de equidad e inclusión. Asimismo, es fundamental medir indicadores de impacto académico, tales como tasas de retención, mejora en el rendimiento y satisfacción estudiantil, a fin de sustentar decisiones sobre su escalamiento institucional. La adopción de modelos predictivos en tutorías académicas debe concebirse como un proceso gradual y sostenible, que trascienda la mera implementación tecnológica. Es indispensable consolidar una cultura institucional de innovación educativa, fortaleciendo las capacidades del personal académico en el uso de analítica de datos y fomentando la participación activa de estudiantes y tutores. Solo así será posible garantizar que la tecnología se convierta en un medio para potenciar el aprendizaje y no en un fin en sí mismo.

### **Referencias Bibliográficas**

- Ausubel, N. (1983). Aprendizaje por American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). APA Publishing.
- Arias, F. (2012). El proyecto de investigación: Introducción a la metodología científica (6.<sup>a</sup> ed.). Episteme.
- Bardin, L. (2011). Análisis de contenido. Akal.
- Beauchamp, T., & Childress, J. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.
- Constitución de la República del Ecuador. (2008). Registro Oficial No. 449.
- Emanuel, E. (2016). Reforming the health care system. *New England Journal of Medicine*,

- 375(12), 1201–1204.  
<https://doi.org/10.1056/NEJMsb1609873>
- Flick, U. (2018). *An introduction to qualitative research* (5th ed.). SAGE Publications.
- Hernández, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación (6.<sup>a</sup> ed.). McGraw-Hill Education.
- Instituto Nacional de Estadística y Censos. (2022). Tecnologías de la información y comunicación (TIC) en los hogares. INEC.  
<https://www.ecuadorencifras.gob.ec>
- Kaplan, A., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68.  
<https://doi.org/10.1016/j.bushor.2009.09.003>
- Kotler, P., Kartajaya, H., & Setiawan, I. (2017). Marketing 4.0: Moving from traditional to digital. John Wiley & Sons.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.
- Ley Orgánica de Defensa del Consumidor. (2000). Registro Oficial Suplemento No. 116.
- Ley Orgánica de Protección de Datos Personales. (2021). Registro Oficial Suplemento No. 459.
- Ley Orgánica de Salud. (2006). Registro Oficial Suplemento No. 423.
- Lorenzetti, R. (2019). Responsabilidad civil de los médicos (2.<sup>a</sup> ed.). Rubinzal-Culzoni Editores.
- Organización Mundial de la Salud. (2020). Ethics and governance of artificial intelligence for health. OMS.  
<https://www.who.int>
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.
- Rodríguez, P., & García, M. (2021). Protección de datos personales y derecho a la intimidad en el ámbito sanitario. *Revista Iberoamericana de Derecho Informático*, 12(2), 45–67.
- Ventola, C. (2014). Social media and health care professionals: Benefits, risks, and best practices. *Pharmacy and Therapeutics*, 39(7), 491–520.



Esta obra está bajo una licencia de Creative Commons Reconocimiento-No Comercial 4.0 Internacional. Copyright © Washington Raúl Fierro Saltos, Elizabeth Alexandra Veloz Segura, Amalín Ladaysé Mayorga Albán y Fabián Rivera.

