DOI: https://doi.org/10.26620/uniminuto.inventum.21.42.2026 ISSN 1909-2520 (Impreso), ISSN 2590-8219 (En línea)

Predicción temprana de deserción universitaria: revisión sistemática de modelos predictivos

Early prediction of university dropout: A systematic review of predictive models

Previsão precoce do abandono universitário: revisão sistemática de modelos preditivos

A. F. Flores-Lira, M. E. Meda-Campaña y J. Arámburo-Lizárraga.

***Resumen*—**La deserción en la educación superior representa un desafío crítico con implicaciones académicas, sociales y económicas. En México, las tasas de abandono oscilan entre el 8 % y el 20 %, afectando programas como la Licenciatura en Tecnologías de la Información de la Universidad de Guadalajara. Este artículo presenta una revisión sistemática de modelos predictivos aplicados a la identificación temprana del abandono estudiantil, mediante el enfoque PRISMA y el análisis de 15 estudios relevantes en bases de datos como Scopus, Web of Science y ERIC. Los hallazgos destacan el uso de técnicas como árboles de decisión, redes neuronales y regresión logística. Sin embargo, persisten desafíos como la limitada integración de variables psicosociales y la escasa aplicación oportuna de los modelos predictivos durante la trayectoria escolar. Se propone un marco metodológico integral para el desarrollo de modelos predictivos eficaces y éticos, con énfasis en enfoques interdisciplinarios orientados a reducir la deserción en la educación superior, particularmente en contextos públicos como el de México.

***Palabras clave*—**Deserción universitaria, modelos predictivos, aprendizaje automático, regresión logística, árboles de decisión, minería de datos, revisión sistemática.

***Abstract*—**Student dropout in higher education is a critical challenge with academic, social, and economic implications. In Mexico, dropout rates range from 8% to 20%, affecting programs such as the Bachelor's Degree in Information Technologies at the University of Guadalajara. This article presents a systematic review of predictive models applied to the early identification of student dropout, using the PRISMA approach and analyzing 15 relevant studies from databases such as Scopus, Web of Science, and ERIC. Findings highlight the frequent use of techniques such as decision trees, neural networks, and logistic regression. However, challenges remain, such as the limited integration of psychosocial variables and the scarce timely application of predictive models throughout the students' academic trajectory. This study proposes a comprehensive methodological framework for developing effective and ethical predictive models, emphasizing interdisciplinary approaches aimed at reducing dropout in higher education, particularly in public institutions in Mexico.

***Keywords*—**University dropout, predictive models, machine learning, logistic regression, decision trees, data mining, systematic review.

***Resumo*—**A evasão no ensino superior representa um desafio crítico com implicações acadêmicas, sociais e econômicas. No México, as taxas de evasão variam entre 8% e 20%, afetando cursos como o Bacharelado em Tecnologias da Informação da Universidade de Guadalajara. Este artigo apresenta uma revisão sistemática de modelos preditivos aplicados à identificação precoce da evasão estudantil, utilizando a abordagem PRISMA e a análise de 15 estudos relevantes em bases como Scopus, Web of Science e ERIC. Os resultados destacam o uso frequente de técnicas como árvores de decisão, redes neurais e regressão logística. No entanto, persistem desafios, como a limitada integração de variáveis psicossociais e a pouca aplicação oportuna dos modelos preditivos ao longo da trajetória acadêmica dos estudantes. Este estudo propõe uma estrutura metodológica abrangente para o desenvolvimento de modelos preditivos eficazes e éticos, com ênfase em abordagens interdisciplinares voltadas à redução da evasão no ensino superior, especialmente em contextos públicos como o do México.

[1] Producto derivado del proyecto de investigación "Modelo predictivo para identificar estudiantes en riesgo de deserción temprana en educación superior", apoyado por el Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara a través del Doctorado en Tecnologías de Información.

A. F. Flores-Lira, Universidad Guadalajara, Zapopan, México, email: favian.flores@academicos.udg.mx.

M. E. Meda-Campaña, Universidad Guadalajara, Zapopan, México, email:elena.meda@academicos.udg.mx

J. Arámburo-Lizárraga, Universidad Guadalajara, Zapopan, México, email: jaramburo@cucea.udg.mx

Cómo citar este artículo: Flores-Lira, A. F., Meda-Campaña, M. E., y Arámburo-Lizárraga, J. Predicción temprana de deserción universitaria: revisión sistemática de modelos predictivos, Entre Ciencia e Ingeniería, vol. 20, no. 37.

Palavras-chave— Evasão universitária, modelos preditivos, aprendizagem de máquina, regressão logística, árvores de decisão, mineração de dados, revisão sistemática.

I. Introducción

La deserción en la educación superior es una problemática compleja que afecta tanto a las instituciones educativas como a los propios estudiantes. Sus consecuencias abarcan desde pérdidas económicas hasta impactos en la trayectoria académica y profesional de quienes interrumpen sus estudios. Diversos estudios han documentado este fenómeno, los cuales coinciden en que las causas suelen ser multifactoriales, incluyendo aspectos académicos, socioemocionales, económicos y del entorno institucional [3], [5], [19], [20].

En México, las tasas de abandono en programas universitarios oscilan entre el 8 % y el 20 % [25], [26]. No obstante, estos promedios pueden ser más elevados en programas específicos, particularmente en áreas como tecnologías de la información, donde el ritmo de actualización y las exigencias académicas requieren un alto nivel de adaptación por parte de la comunidad estudiantil. Un análisis reciente de la Licenciatura en Tecnologías de la Información en el Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara reportó una tasa de deserción temprana del 24.8 % durante los primeros dos años. Este dato no solo supera el promedio nacional, sino que también refleja los retos particulares que enfrentan los programas de formación en este campo.

Esta realidad refuerza la necesidad urgente de implementar estrategias que permitan anticipar y atender el abandono desde los primeros semestres. Si bien existen enfoques cualitativos y herramientas diagnósticas que ayudan a comprender las causas del abandono —como los aplicados en encuestas nacionales [11]— estos suelen utilizarse cuando el riesgo ya es evidente. En cambio, los modelos predictivos permiten identificar patrones de riesgo desde etapas tempranas, a partir de datos disponibles en los registros institucionales. Esta capacidad de anticipación es lo que les otorga un valor añadido en contextos como el de la LTIN, donde la deserción ocurre de manera acelerada. Además, su aplicación facilita intervenciones personalizadas y oportunas, antes de que la decisión de abandonar los estudios se concrete [1], [4], [5].

El objetivo de este artículo es presentar una revisión sistemática de investigaciones recientes sobre modelos predictivos aplicados a la deserción estudiantil en educación superior. El análisis toma en cuenta las metodologías empleadas, las variables más comunes y los niveles de precisión que se reportan en la literatura. Entre los modelos que se han utilizado con mayor frecuencia destacan la regresión logística, los árboles de decisión, Random Forest y las redes neuronales, incluyendo variantes como la regresión iterativa [16], los cuales han mostrado distintos niveles de efectividad según el contexto. A partir de esta revisión, se propone un marco metodológico que sirva de base para el desarrollo de un modelo predictivo más preciso y pertinente, enfocado en la detección temprana del abandono en la Licenciatura en Tecnologías de la Información.

II. Contexto Institucional

En los últimos años, el interés por comprender y prevenir la deserción estudiantil ha impulsado el desarrollo de diversos enfoques analíticos, especialmente aquellos basados en técnicas de modelado predictivo. La evolución de herramientas computacionales y la disponibilidad de grandes volúmenes de datos han permitido que áreas como el aprendizaje automático y la minería de datos se integren de manera progresiva en el ámbito educativo. Este cambio responde a la necesidad de detectar patrones tempranos de abandono, superando las limitaciones de métodos tradicionales centrados exclusivamente en encuestas o percepciones cualitativas.

En este escenario, los modelos predictivos se han posicionado como instrumentos claves para anticipar el riesgo de deserción y diseñar estrategias de intervención personalizadas y eficientes. A continuación, se presenta una revisión sistemática de los principales enfoques y modelos aplicados en la literatura reciente, se considera como punto de partida los hallazgos preliminares obtenidos en el análisis institucional de la LTIN.

Como parte de la investigación, se realizó un análisis preliminar de los cohortes 2019B a 2022B de la LTIN. Este periodo fue seleccionado por corresponder a los ciclos escolares más recientes con datos completos, disponibles y homogéneos en los registros institucionales, lo que permite obtener una visión actualizada y confiable de las tendencias de abandono.

El estudio consideró una muestra de 794 estudiantes de nuevo ingreso. En la Tabla 1 se muestra que, de los 283 estudiantes que desertaron durante los primeros cuatro semestres, 197 lo hicieron durante el primer año, es decir, representan un 69.6 % del abandono concentrado al inicio de la trayectoria académica. Así, respecto al total de estudiantes de nuevo ingreso, esto equivale a una tasa de deserción temprana del 24.8 %. Este hallazgo refuerza la necesidad de contar con herramientas que permitan anticipar el riesgo desde las etapas iniciales, por lo que se justifica la revisión sistemática de modelos predictivos presentada en este artículo.

TABLA 1.
TASA DE DESERCIÓN POR SEMESTRE EN LA LTIN-CUCEA (2019B–2022B)

Ciclos	Ingreso	Desertores				Total
		Sem. 1	Sem. 2	Sem. 3	Sem. 4	
2019A	84	10	7	1	2	20
2019B	85	14	5	2	3	24
2020A	88	16	6	9	6	37
2020B	102	12	16	7	14	49
2021A	105	20	9	3	4	36
2021B	105	20	14	5	6	45
2022A	111	17	10	7	2	36
2022B	114	15	6	5	5	31

DOI: https://doi.org/10.26620/uniminuto.inventum.21.42.2026

Totales	794	124	73	39	47	283
		Primer año		**Segundo año**		
		197		86		

A partir de los datos anteriores, se elaboraron dos representaciones gráficas para facilitar la interpretación visual de las tendencias de abandono. La Figura 1 muestra la deserción total agrupada por año académico (primer y segundo año), mientras que la Figura 2 detalla los casos por semestre. Ambas figuras muestran con claridad que el mayor número de abandonos ocurre durante el primer año, donde se observa que el primer semestre representa el de mayor concentración de deserción.

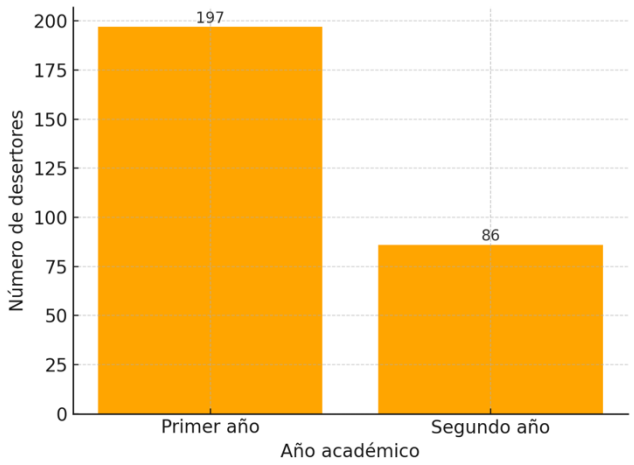

Fig 1. Deserción por año académico LTIN-CUCEA (2019B–2022B)

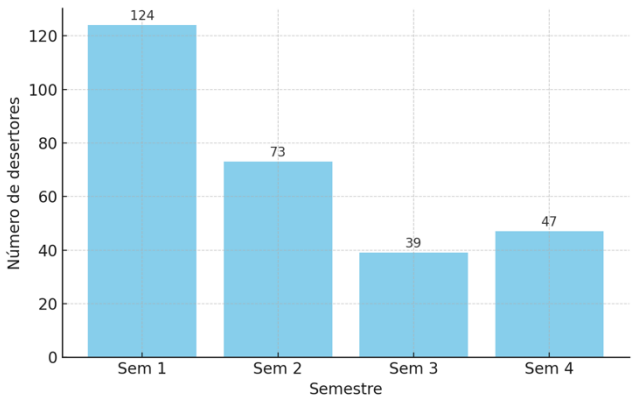

Fig. 2. Deserción acumulada por semestre LTIN-CUCEA (2019B–2022B)

III. METODOLOGÍA DE REVISIÓN SISTEMÁTICA

Para la elaboración de la revisión sistemática sobre modelos predictivos aplicados a la deserción universitaria, se adoptó el enfoque PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [18], una metodología ampliamente utilizada en revisiones sistemáticas para garantizar transparencia, trazabilidad y rigor en el proceso de búsqueda, selección y análisis de la literatura científica. Este enfoque permite documentar cada etapa, desde la identificación de estudios hasta su selección final, asegurando que el análisis sea transparente, reproducible y fundamentado en criterios definidos previamente.

Fuentes de información y estrategias de búsqueda

Se realizó una búsqueda documental entre febrero y abril de 2025 en las bases de datos Scopus, Web of Science (Clarivate Analytics), IEEE Xplore y ERIC. Estas plataformas fueron seleccionadas por su alta indexación de artículos arbitrados en los campos de la educación, informática, ciencia de datos y aprendizaje automático.

Para la estrategia de búsqueda, se utilizaron combinaciones de palabras clave en inglés como "university dropout", "student attrition", "predictive model", "machine learning", "logistic regression" y "higher education". Asimismo, se aplicaron operadores booleanos (AND, OR) para ampliar los resultados y filtros por idioma (español e inglés), tipo de publicación (artículo científico) y periodo de publicación (entre 2015 y 2024).

Criterios de inclusión y exclusión

Para la búsqueda de los trabajos se establecieron los siguientes criterios:

Criterios de inclusión:

1. Estudios empíricos que implementan o evalúan modelos predictivos de deserción universitaria.
2. Artículos publicados en revistas arbitradas.
3. Estudios con acceso completo al texto.
4. Publicaciones en español o inglés.

Criterios de exclusión:

1. Trabajos enfocados en niveles educativos distintos a la educación superior.
2. Estudios puramente teóricos o sin evidencia empírica.
3. Publicaciones duplicadas o con información incompleta.

Proceso de selección

La búsqueda inicial arrojó un total de 194 artículos. Tras la eliminación de duplicados y una primera revisión de títulos y resúmenes, se seleccionaron 36 estudios para revisión a texto completo. Finalmente, 15 artículos cumplieron con los criterios establecidos y fueron incluidos en el análisis sistemático. El proceso completo de selección se resume en la Figura 3, conforme al diagrama de flujo propuesto por la metodología PRISMA.

DOI: https://doi.org/10.26620/uniminuto.inventum.21.42.2026

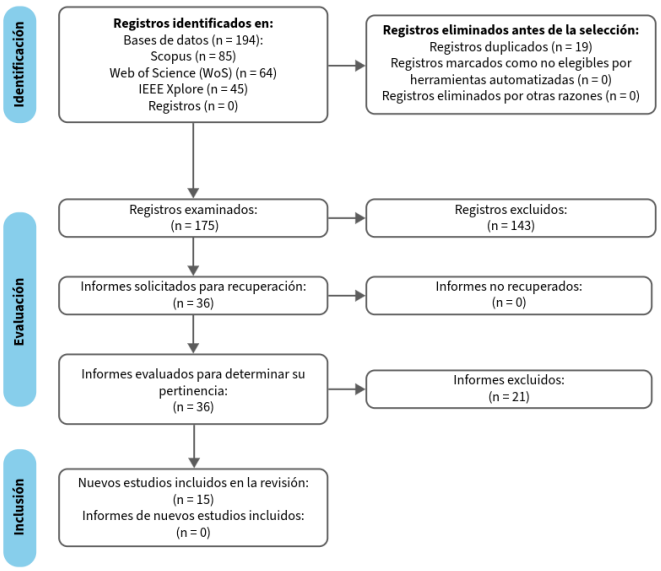

Fig 3. Diagrama de flujo PRISMA que resume el proceso de identificación, selección e inclusión de estudios en la revisión sistemática.

En la Tabla 2 se presentan los 15 estudios que conforman la base del análisis, mismos que fueron seleccionados tras la aplicación de los criterios metodológicos PRISMA. Estos trabajos abordan la predicción de la deserción en educación superior mediante distintos enfoques de aprendizaje automático y estadístico. Para facilitar su lectura, la tabla incluye el modelo predictivo utilizado en cada caso, así como una breve descripción de su aporte principal al campo. Los identificadores DOI de cada estudio se encuentran disponibles en la sección de referencias.

TABLA 2.
ESTUDIOS INCLUIDOS EN LA REVISIÓN SISTEMÁTICA

Autor(es) y Año	Título del estudio	Modelo predictivo utilizado	Aporte principal
Aulck, L., Aras, R., Li, L., L'Heureux, C., Lu, P., y West, J. (2017)	Stem-ming the Tide: Predicting STEM attrition using student transcript data	Regresión logística	Uso de datos de transcripción para predecir abandono en carreras STEM (Science, Technology, Engineering, and Mathematics).
Brdesee, H. S., Alsaggaf, W., Aljohani, N., y Hassan, S.-U. (2022)	Predictive Model Using a Machine Learning Approach for Enhancing the Retention Rate of Students At-Risk	Random Forest, SVM (Support Vector Machines)	Mejora de la retención con enfoque supervisado para estudiantes en riesgo
Bulut, O., Wongvorachan, T., He, S., y Lee, S. (2024)	Enhancing high-school dropout identification: a collaborative approach	Enfoque colaborativo Human-ML (Human integrating human and machine insights | Machine Learning)	Integración de percepciones humanas con modelos automáticos
Cornejo Sifuentes, S. G., Vega Pérez, L. G., Naranjo Cantabrana, M. G., Osúa Acosta, I. F., Avila Santana, J. F., y Ramírez Santamaría, M. A. (2023)	Modelo predictivo de la deserción escolar en educación superior: una aproximación desde la minería de datos utilizando la metodología CRISP-DM	Árboles de decisión	Aplicación de CRISP-DM (Cross-Industry Standard Process for Data Mining) en minería de datos educativa
Del Bonifro, F., Gabbrielli, M., Lisanti, G., y Zingaro, S. P. (2020)	Student Dropout Prediction	Redes neuronales	Predicción con redes neuronales usando características académicas
Delgado Celis, M. D., Navarrete Sánchez, M. E., Rosales Escobar, M. L., Félix Jácquez, R. H., y Cruz Rentería, M. M. (2022)	El problema de la deserción en educación superior y cómo predecirla	Regresión logística	Aplicación de modelo explicativo y predictivo en contexto mexicano
Henríquez Cabezas, N., y Vargas Escobar, D. (2022)	Modelos predictivos de rendimiento y deserción académica en estudiantes de primer año de una universidad pública chilena	Regresión logística	Predicción en estudiantes de primer año en universidad pública chilena
Hoyos Osorio, J. K., y Daza Santacoloma, G. (2023)	Modelo predictivo para identificar estudiantes universitarios con alto grado de deserción	Random Forest	Predicción temprana basada en variables académicas y socioeconómicas
Karimi-Haghighi, M., Castillo, C., Hernández-Leo, D., y Moreno Oliver, V. (2021)	Predicting Early Dropout: Calibration and Algorithmic Fairness Considerations	Modelos calibrados, equidad algorítmica	Consideraciones éticas y de equidad en predicción automatizada
Kim, H. (2023)	Predicting College Student Dropouts with Machine Learning	Árboles de decisión, regresión logística	Comparación de modelos supervisados con alta precisión
Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., y Addison, K. L. (2015)	A machine learning framework to identify students at risk of adverse academic outcomes	Random Forest	Marco general de predicción en contextos educativos diversos

Núñez Villalobos, D. (2025)	Modelo predictivo basado en aprendizaje automático para la retención estudiantil en educación superior	Regresión logística, árboles de decisión	Modelo enfocado en retención en educación superior en Perú
Okoye, K., Nganji, J. T., Escamilla, J., y Hosseini, S. (2024)	Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education	ML Ensemble (Machine Learning Ensemble)	Predicción de retención y graduación usando combinación de algoritmos
Sulak, S. A., y Koklu, N. (2024)	Predicting Student Dropout Using Machine Learning Algorithms	Árboles de decisión, Naive Bayes	Evaluación comparativa de algoritmos ML (Machine Learning) clásicos
Zhang, L., y Rangwala, H. (2018)	Early Identification of At-Risk Students Using Iterative Logistic Regression	Regresión logística iterativa	Identificación de riesgo temprano en estudiantes universitarios

IV. Resultados de la Revisión

Al revisar los 15 estudios seleccionados (ver Tabla 2), se observa una amplia variedad de enfoques para abordar la predicción de la deserción en la educación superior. Aunque todos aplican técnicas cuantitativas, algunos se apoyan en métodos estadísticos tradicionales como la regresión logística, mientras que otros optan por algoritmos más recientes como Random Forest, redes neuronales o modelos híbridos más complejos. En particular, estudios como los de Cornejo Sifuentes et al. [23] y Costa et al. [6] conbinaron árboles de decisión con la metodología CRISP-DM y técnicas de minería de datos educativa, logrando resultados positivos en universidades de América Latina.

La regresión logística continúa siendo uno de los enfoques más utilizados, especialmente en investigaciones realizadas en América Latina [5], [17], posiblemente porque resulta fácil de implementar y permite interpretar los resultados sin demasiada complejidad. De los 15 estudios revisados en este trabajo 6 de ellos utilizaron este enfoque, ya sea como técnica principal o como parte de una comparación con otros algoritmos. Sin embargo, a pesar de su popularidad, investigaciones más recientes —como las de Kim [8], Karimi-Haghighi et al. [19] o Sulak y Koklu [22] — muestran que modelos más sofisticados, como Random Forest o redes neuronales, tienden a ofrecer mejores niveles de precisión en determinados contextos.

En relación con las variables empleadas en los modelos, la mayoría de los estudios coinciden en analizar datos académicos como el promedio general, materias reprobadas o tipo de ingreso, como se observa en trabajos de Delgado Celis et al. [17], Núñez Villalobos [5], Hoyos Osorio y Daza Santacoloma [12] o Aulck et al. [15]. También es común que se incluyan factores demográficos y socioeconómicos, como edad, género o nivel de ingresos, lo cual puede verse en investigaciones como las de Henríquez Cabezas y Vargas Escobar [20] o Brdesee et al. [10]. Algunos estudios, como los de Karimi-Haghighi et al. [19] y Okoye et al. [13], incluso incorporan elementos del entorno institucional o el uso de plataformas virtuales, lo que permite enriquecer la predicción con más dimensiones del comportamiento estudiantil.

Es importante destacar que una parte significativa de estos estudios se ha desarrollado en países como México, Colombia, Perú y Ecuador. Esto refleja no solo un creciente interés por entender y prevenir la deserción, sino también una necesidad práctica en muchas universidades públicas de la región. A pesar de estos avances, todavía es poco común encontrar modelos predictivos que estén integrados de manera activa en los sistemas de gestión institucional o que se apliquen durante el ciclo escolar para anticipar y prevenir el abandono en tiempo oportuno.

En términos de evaluación, algunos de los estudios revisados reportan métricas como precisión o AUC, lo que permite valorar de manera más objetiva la efectividad de los modelos aplicados. La precisión refleja el porcentaje de aciertos en las predicciones, mientras que el AUC (Área Bajo la Curva ROC) mide qué tan bien el modelo distingue entre estudiantes que abandonan y los que continúan. Estas métricas aparecen claramente en los trabajos de Del Bonifro et al. [7] y Brdesee et al. [10], donde se utilizan para comparar el desempeño de los modelos probados. A pesar de ello, algunos estudios no explican con suficiente detalle su proceso de validación o las características de sus muestras, lo que dificulta replicar o contrastar sus resultados de manera rigurosa.

En general, los estudios analizados muestran avances importantes en el uso de técnicas predictivas para identificar a estudiantes en riesgo de abandonar sus estudios. No obstante, también dejan ver que aún existen oportunidades para mejorar la integración institucional de estos modelos y para trabajar en metodologías más estandarizadas y comparables entre sí.

V. Análisis de las Variables más Significativas en los Estudios Revisados

Uno de los elementos clave en los modelos predictivos revisados es el tipo de variables que se consideran para anticipar el abandono estudiantil. Aunque todos los estudios analizados incorporan al menos un conjunto de variables académicas, existen diferencias importantes en su frecuencia de uso y relevancia estadística.

En la revisión de los 15 artículos seleccionados, se identificó que las variables académicas son las más utilizadas y con mayor poder predictivo, especialmente el promedio general, las materias reprobadas y los créditos cursados o aprobados, las cuales se encuentran presentes en al menos 7 estudios. Estas

variables aparecen de forma destacada en los trabajos de Núñez Villalobos [5], Del Bonifro et al. [7], Kim [8], y Delgado Celis et al. [17], entre otros.

Por su parte, las variables demográficas como género y edad también muestran una presencia significativa. El género fue incluido en 10 estudios, como en los de Brdesee et al. [10], Hoyos Osorio y Daza Santacoloma [12], y Karimi-Haghighi et al. [19], mientras que la edad fue considerada en 8 estudios, usualmente asociada a estudiantes mayores de 25 años o que combinan estudio y trabajo.

Respecto a las condiciones económicas, se unificaron las variables "nivel socioeconómico" e "ingreso familiar" en una categoría más amplia denominada "condición económica", presente en 9 estudios *(Nota: esta categoría agrupa tanto estudios que reportan el ingreso familiar directo como aquellos que infieren el nivel socioeconómico mediante otros indicadores, como ocupación de los padres, tipo de vivienda o acceso a apoyos económicos)*. En trabajos como los de Núñez Villalobos [5], Delgado Celis et al. [17] y Cornejo Sifuentes et al. [23], se reporta que un entorno económico desfavorable puede aumentar considerablemente el riesgo de abandono, especialmente en contextos de vulnerabilidad social o instituciones públicas.

De uso menos común son las variables institucionales y psicosociales, que incluyen desde la utilización de plataformas virtuales hasta la percepción del acompañamiento académico o el bienestar emocional. Estos elementos aparecen en estudios como los de Karimi-Haghighi et al. [19], Okoye et al. [13] y Sulak y Koklu [22], donde se exploran también aspectos como la equidad algorítmica y la motivación del estudiante, aportando una mirada más integral del fenómeno de deserción.

La siguiente tabla muestra un resumen de las variables más comunes y su relevancia observada en los estudios revisados.

TABLA 3.

VARIABLES MÁS UTILIZADAS EN LOS MODELOS PREDICTIVOS DE DESERCIÓN REVISADOS

Variable	Frecuencia de uso (15 estudios)	Tipo	Estudios donde destaca
Promedio general	10	Académica	[5], [7], [8], [9], [10], [13], [15], [16], [20], [21]
Materias Reprobadas	10	Académica	[5], [7], [10], [12], [13], [15], [16], [17], [20], [23]
Género	10	Demográfica	[5], [7], [8], [10], [12], [13], [15], [17], [19], [23]
Condición económica (nivel socioeconómico o ingreso familiar)	9	Socioeconómica	[5], [7], [8], [10], [15], [17], [19], [20], [23]
Edad	8	Demográfica	[5], [7], [8], [10], [12], [13], [15], [19]
Créditos cursados o aprobados	7	Académica	[5], [7], [8], [12], [13], [15], [17]
Uso de plataformas institucionales	3	Institucional	[13], [19], [22]
Apoyo Familiar o emocional	3	Psicosocial	[13], [19], [22]

Estos hallazgos refuerzan la importancia de combinar variables académicas tradicionales con otras de tipo contextual o emocional, lo cual puede resultar clave para lograr modelos predictivos completos, humanos y útiles en la práctica institucional.

VI. DISCUSIÓN Y RECOMENDACIONES

La revisión sistemática deja ver que, aunque existe un interés creciente por aplicar modelos predictivos a la problemática de la deserción en la educación superior, aún persisten varios retos tanto a nivel metodológico como en su implementación práctica. En este sentido, la presente revisión se apoyó en la guía PRISMA [18], lo cual permitió garantizar una búsqueda estructurada, transparente y replicable, siguiendo estándares internacionales para revisiones sistemáticas. Este enfoque metodológico contribuye a reforzar la validez de los hallazgos presentados y a facilitar comparaciones con futuros estudios similares.

Por un lado, si bien hay un avance en el uso de algoritmos más sofisticados como Random Forest, redes neuronales o modelos híbridos, la falta de criterios estandarizados para seleccionar variables o evaluar modelos representa un área de oportunidad para presentar resultados que aporten a una selección de criterios adecuados y eficientes. Cada estudio utiliza diferentes conjuntos de datos, variables y formas de validación, lo que dificulta hacer comparaciones directas o extraer conclusiones generalizables.

Por otro lado, aunque muchos modelos reportan buenos niveles de precisión, esto no garantiza su aplicación real en contextos institucionales. Son pocos los trabajos que mencionan si sus modelos fueron adoptados o probados dentro de universidades, lo cual sugiere una brecha entre el desarrollo académico y su uso práctico. De hecho, en muchos casos no se aborda cómo estos modelos podrían integrarse a los sistemas de información que ya operan en las instituciones. Además, recientes propuestas sobre inteligencia artificial explicativa [24] destacan la importancia de incorporar mecanismos que justifiquen las decisiones del modelo, aspecto clave para su aceptación en entornos educativos reales.

También se identificó que la mayoría de los estudios se centra en variables académicas y demográficas, mientras que aspectos como el bienestar emocional, la motivación o las redes de apoyo social todavía se consideran poco, a pesar de su

DOI: https://doi.org/10.26620/uniminuto.inventum.21.42.2026

posible impacto en la permanencia estudiantil. Esto abre una línea interesante para estudios futuros que integren datos cualitativos o variables psicosociales.

En cuanto a las regiones estudiadas, es positivo ver que en América Latina —y especialmente en México— comienza a desarrollarse una base importante de investigaciones sobre este tema. No obstante, aún son escasos los estudios longitudinales o con datos de múltiples cohortes que permitan hacer análisis robustos y predecir tendencias a mediano o largo plazo.

Como recomendación general, se considera pertinente fomentar la colaboración interinstitucional para generar bases de datos compartidas y fomentar líneas de investigación colaborativa que permitan mejorar la calidad y aplicabilidad de los modelos. Asimismo, sería útil que las universidades comiencen a explorar formas de integrar estos modelos a sus sistemas internos de alerta temprana, no solo como herramienta de análisis, sino como mecanismo activo de prevención.

Además, si bien los modelos predictivos permiten identificar a los estudiantes en riesgo, comprender sus trayectorias estudiantiles —es decir, los recorridos académicos que construyen a lo largo del tiempo, con posibles pausas, cambios o abandonos— ofrece una perspectiva más amplia sobre el fenómeno. Considerar estas trayectorias permite no solo predecir el abandono, sino también entender en qué momento y bajo qué condiciones ocurre, lo cual resulta clave para diseñar intervenciones más oportunas y eficaces.

Finalmente, para fortalecer el campo, sería ideal contar con una guía metodológica común que oriente a los investigadores en el diseño, evaluación y publicación de modelos predictivos de deserción, facilitando así la comparación entre estudios y el avance conjunto del conocimiento en esta área.

VII. Conclusiones

La presente revisión sistemática confirma que el uso de modelos predictivos en la educación superior representa una estrategia prometedora para anticipar la deserción estudiantil, especialmente durante los primeros semestres. A través del análisis de 15 estudios recientes, se identificó una evolución significativa en las técnicas empleadas, desde métodos estadísticos tradicionales como la regresión logística hasta enfoques más complejos basados en aprendizaje automático, como redes neuronales o Random Forest.

Si bien la precisión de los modelos ha mejorado considerablemente, persisten desafíos importantes. Entre ellos destacan la limitada integración de variables psicosociales, la escasa estandarización metodológica y la débil conexión con los sistemas institucionales de alerta temprana. Estos factores dificultan la aplicación práctica de los modelos en contextos reales y ponen de manifiesto la necesidad de contar con datos longitudinales y colaboraciones interinstitucionales que permitan desarrollar soluciones óptimas y adaptables.

En el contexto específico del CUCEA, los datos históricos de la Licenciatura en Tecnologías de la Información evidencian que cerca del 25 % del abandono estudiantil ocurre en el primer año. Esta realidad no solo valida los hallazgos reportados en la literatura internacional, sino que también refuerza la urgencia de adoptar modelos predictivos contextualizados, capaces de considerar las particularidades académicas, sociales y culturales de cada institución.

Dicha situación refleja una problemática común a muchas universidades públicas en México, donde el abandono temprano suele estar vinculado a una combinación de factores que trascienden el rendimiento académico. Por ello, resulta indispensable avanzar hacia un enfoque integral que contemple también dimensiones socioemocionales, económicas y de integración institucional, a fin de intervenir de manera más oportuna y efectiva. Como parte de las estrategias sugeridas, se propone el desarrollo de un sistema automatizado de alerta temprana que integre modelos predictivos con las bases de datos académicas existentes. Este sistema permitiría identificar a estudiantes en riesgo desde su ingreso, para generar recomendaciones personalizadas para su acompañamiento.

Complementariamente, diversos estudios sugieren que estrategias como talleres de inducción más sólidos, tutorías académicas proactivas y mecanismos de seguimiento emocional (como encuestas periódicas o espacios de mentoría estudiantil) pueden ser efectivas para reducir la deserción. Instituciones en países como México, Colombia y Chile ya han comenzado a implementar estas acciones dentro de políticas institucionales enfocadas en la permanencia estudiantil [9], [14], [20]. Estas medidas, cuando se articulan adecuadamente con modelos predictivos, se convierten en herramientas clave para intervenir de manera oportuna y personalizada, favoreciendo la permanencia en la educación superior.

En suma, avanzar en la predicción temprana de la deserción universitaria no implica únicamente mejorar los algoritmos, sino también comprender las trayectorias estudiantiles en toda su complejidad. Estas trayectorias representan los diversos recorridos que los estudiantes construyen a lo largo de su paso por la universidad, incluyendo avances, interrupciones, cambios o abandonos. Comprender estos procesos permite identificar en qué momento y bajo qué condiciones ocurre la deserción, lo cual resulta clave para diseñar intervenciones oportunas y efectivas.

Agradecimientos

Este trabajo se realizó con el respaldo de la comunidad académica y el acompañamiento metodológico recibido en el marco del programa de doctorado. Agradezco profundamente a los docentes y colegas que ofrecieron retroalimentación valiosa durante el proceso de elaboración. Este esfuerzo es el reflejo de muchas manos invisibles que, a través de ideas, datos y preguntas, contribuyen a construir una educación superior más

equitativa y con menores índices de deserción.

REFERENCIAS

[1] A. Ramesh, D. Goldwasser, B. Huang, H. Daumë III, y L. Getoor, "Learning latent engagement patterns of students in online courses," en Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-14), vol. 2, pp. 1272–1278, 2014.

[2] A. Villar y C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," *Discover Artificial Intelligence*, vol. 4, art. no. 2, 2024. [En línea]. Disponible en: https://doi.org/10.1007/s44163-023-00079-z

[3] B. R. Villegas y L. A. Núñez Lira, "Factores asociados a la deserción estudiantil en el ámbito universitario: Una revisión sistemática 2018–2023," RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo, vol. 14, no. 28, 2024. [En línea]. Disponible en: https://doi.org/10.23913/ride.v14i28.1923

[4] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," Comput. Educ., vol. 131, pp. 22–32, 2019. [En línea]. Disponible en: https://doi.org/10.1016/j.compedu.2018.12.006

[5] D. Núñez Villalobos, "Modelo predictivo basado en aprendizaje automático para la retención estudiantil en educación superior," European Public & Social Innovation Review, vol. 10, 2025. [En línea]. Disponible en: https://doi.org/10.31637/epsir-2025-1307

[6] E. Costa, B. Fonseca, M. Santana, and F. de Araújo, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure," Comput. Human Behav., vol. 73, pp. 247–256, 2017. [En línea]. Disponible en: https://doi.org/10.1016/j.chb.2017.01.047

[7] F. Del Bonifro, M. Gabbrielli, G. Lisanti y S. P. Zingaro, "Student Dropout Prediction," en Artificial Intelligence in Education. AIED 2020, Lecture Notes in Artificial Intelligence, vol. 12163, Springer, 2020, pp. 129–140. [En línea]. Disponible en: https://doi.org/10.1007/978-3-030-52237-7_11

[8] H. Kim, "Predicting College Student Dropouts with Machine Learning," Research Archive of Rising Scholars, preprint, 28 de octubre de 2023. [En línea]. Disponible en: https://doi.org/10.58445/rars.653

[9] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani y K. L. Addison, "A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes," en Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15), Sydney, Australia, 2015, pp. 1909–1918. [En línea]. Disponible en: https://doi.org/10.1145/2783258.2788620

[10] H. S. Brdesee, W. Alsaggaf, N. Aljohani y S.-U. Hassan, "Predictive Model Using a Machine Learning Approach for Enhancing the Retention Rate of Students At-Risk," International Journal on Semantic Web and Information Systems, vol. 18, núm. 1, pp. 1–21, 2022. [En línea]. Disponible en: https://doi.org/10.4018/IJSWIS.299859

[11] Instituto Nacional de Estadística y Geografía (INEGI), "Tasa de abandono escolar en educación superior," 2023–2024. [En línea]. Disponible en: https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=9171df60-8e9e-4417-932e-9b80593216ee

[12] J. K. Hoyos Osorio y G. Daza Santacoloma, "Modelo predictivo para identificar estudiantes universitarios con alto grado de deserción," Revista Electrónica de Investigación Educativa, vol. 25, e13, pp. 1–10, 2023. [En línea]. Disponible en: https://redie.uabc.mx/redie/article/view/5398

[13] K. Okoye, J. T. Nganji, J. Escamilla y S. Hosseini, "Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education," Computers and Education: Artificial Intelligence, vol. 6, art. no. 100205, 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.caeai.2024.100205

[14] K. Rivera Vergaray, "Modelo predictivo para la detección temprana de estudiantes con alto riesgo de deserción académica," Innov. y Softw., vol. 2, no. 2, pp. 6–13, 2021. [En línea]. Disponible en: https://www.redalyc.org/articulo.oa?id=673870839001

[15] L. Aulck, R. Aras, L. Li, C. L'Heureux, P. Lu, y J. West, "Stem-ming the Tide: Predicting STEM attrition using student transcript data," arXiv preprint arXiv:1708.09344, 28 de agosto de 2017. [En línea]. Disponible en: https://arxiv.org/abs/1708.09344

[16] L. Zhang y H. Rangwala, "Early Identification of At-Risk Students Using Iterative Logistic Regression," en Artificial Intelligence in Education. AIED 2018, Lecture Notes in Computer Science, vol. 10947, Springer, 2018, pp. 613–626. [En línea]. Disponible en: https://doi.org/10.1007/978-3-319-93843-1_45

[17] M. D. Delgado Celis, M. E. Navarrete Sánchez, M. L. Rosales Escobar, R. H. Félix Jácquez y M. M. Cruz Rentería, "El problema de la deserción en educación superior y cómo predecirla," Ciencia Latina Revista Científica Multidisciplinar, vol. 6, núm. 4, pp. 5303–5327, 2022. [En línea]. Disponible en: https://doi.org/10.37811/cl_rcm.v6i4.3017

[18] M. J. Page, J. E. McKenzie, P. M. Bossuyt, et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, no. 71, 2021. [En línea]. Disponible en: https://doi.org/10.1136/bmj.n71

[19] M. Karimi-Haghighi, C. Castillo, D. Hernández-Leo y V. Moreno Oliver, "Predicting Early Dropout: Calibration and Algorithmic Fairness Considerations," arXiv preprint arXiv:2103.09068, 16 de marzo de 2021. [En línea]. Disponible en: https://arxiv.org/abs/2103.09068

[20] N. Henríquez Cabezas y D. Vargas Escobar, "Modelos predictivos de rendimiento y deserción académica en estudiantes de primer año de una universidad pública chilena," Revista de Estudios y Experiencias en Educación, vol. 21, núm. 45, pp. 299–314, abr. 2022. [En línea]. Disponible en: 10.21703/0718-5162.v21.n45.2022.015

[21] O. Bulut, T. Wongvorachan, S. He y S. Lee, "Enhancing high-school dropout identification: a collaborative approach integrating human and machine insights," Discover Education, vol. 3, art. no. 109, 2024. [En línea]. Disponible en: https://doi.org/10.1007/s44217-024-00209-4

[22] S. A. Sulak y N. Koklu, "Predicting Student Dropout Using Machine Learning Algorithms," Intelligent Methods In Engineering Sciences, vol. 3, núm. 3, pp. 88–97, 2024. [En línea]. Disponible en: https://doi.org/10.58190/imiens.2024.103

[23] S. G. Cornejo Sifuentes, L. G. Vega Pérez, M. G. Naranjo Cantabrana, I. F. Osúa Acosta, J. F. Avila Santana y M. A. Ramírez Santamaría, "Modelo predictivo de la deserción escolar en educación superior: una aproximación desde la minería de datos utilizando la metodología CRISP-DM," Ciencia Latina Revista Científica Multidisciplinar, vol. 7, núm. 5, pp. 7797–7812, 2023. [En línea]. Disponible en: https://doi.org/10.37811/cl_rcm.v7i5.8363

[24] S. Mustofa, Y. R. Emon, S. B. Mamun, S. A. Akhy y M. A. Rahman, "A novel AI-driven model for student dropout risk analysis with explainable AI insights," Computers and Education: Artificial Intelligence, vol. 8, art. no. 100352, 2024. [En línea]. Disponible en: https://doi.org/10.1016/j.caeai.2024.100352

[25] Secretaría de Educación Pública (SEP), "Anexo 6: Reporte de la Encuesta Nacional de Deserción en la Educación Media Superior (ENDEMS)," 2020. [En línea]. Disponible en:

https://educacionmediasuperior.sep.gob.mx/work/models/sems/Resource/10787/1/images/Anexo_6Reporte_de_la_ENDEMS.pdf

[26] Secretaría de Educación Pública (SEP), "Boletín 100: Desciende 8.1% tasa de abandono escolar en educación superior: SEP," 13 de noviembre de 2023. [En línea]. Disponible en: https://www.gob.mx/sep/articulos/boletin-100-desciende-8-1-tasa-de-abandono-escolar-en-educacion-superior-sep?idiom=es



Anastacio Favián Flores Lira. Es Ingeniero en Telemática por la Universidad de Guadalajara (UDG), Centro Universitario del Norte (CUNorte), Colotlán, Jalisco, México (2009). Obtuvo el grado de Maestro en Tecnologías de la Información (2018) y actualmente cursa el Doctorado en Tecnologías de Información en el Centro Universitario de Ciencias Económico Administrativas (CUCEA), por la misma universidad, Zapopan, Jalisco. Su investigación se enfoca en modelos predictivos e inteligencia artificial para identificar estudiantes en riesgo de deserción. Es académico en la UDG, donde coordinó la Licenciatura en Tecnologías de la Información, impulsando la innovación académica y la transformación digital. Cuenta con más de 13 años de experiencia profesional en desarrollo de software, coordinación de equipos multidisciplinarios y diseño de soluciones para los sectores educativo, industrial y gubernamental. Sus áreas de especialización incluyen desarrollo de software, interfaces web y bases de datos. Sus intereses de investigación abarcan analítica educativa, aprendizaje automático, minería de datos y simulación multiagente.
ORCID: https://orcid.org/0000-0001-8088-7024



María Elena Meda Campaña. Es Ingeniera Industrial por el Instituto Tecnológico de Culiacán, Culiacán, México (1996). Obtuvo la Maestría (1998) y el Doctorado (2022) en Ciencias de la Ingeniería Eléctrica, con especialidad en Control Automático, por el Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), Unidad Guadalajara, México. Actualmente es profesora investigadora en el Departamento de Sistemas de Información de la Universidad de Guadalajara. Cuenta con experiencia en dirección y codirección de tesis de posgrado, participación en comités académicos y colaboración en proyectos de investigación aplicada. Es miembro del Sistema Nacional de Investigadores (SNI), nivel 1. Ha publicado artículos en revistas indexadas, participado en congresos nacionales e internacionales, y colabora activamente en la formación de recursos humanos. Sus líneas de investigación incluyen el modelado, la síntesis, la identificación y el control de sistemas de eventos discretos, así como el estudio de sistemas híbridos y el diagnóstico de faltas.
ORCID: https://orcid.org/0000-0001-8684-8296



Jesús Arámburo Lizárraga. Obtuvo el grado de Doctor en Ciencias, en la especialidad en Ingeniería Eléctrica, en el año 2009 por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV), Unidad Guadalajara, México. Actualmente se desempeña como Profesor Investigador en el Centro Universitario de Ciencias Económico Administrativas (CUCEA) de la Universidad de Guadalajara. Las áreas de interés en investigación se centran en el desarrollo de juegos serios y en el diseño de enfoques multidisciplinarios orientados a la integración y el análisis de la tecnología y la educación, así como en el estudio y aplicación de la inteligencia artificial generativa.
ORCID: https://orcid.org/0000-0002-0370-7891