

# INTEGRACIÓN DE VARIABLES CONTEXTUALES, CONDUCTUALES Y SOCIODEMOGRÁFICAS EN LA PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL EN EDUCACIÓN SUPERIOR: UNA REVISIÓN SISTEMÁTICA

## INTEGRATION OF CONTEXTUAL, BEHAVIORAL, AND SOCIODEMOGRAPHIC VARIABLES IN PREDICTING STUDENT DROPOUT IN HIGHER EDUCATION: A SYSTEMATIC REVIEW

**Tipo de Publicación:** Artículo Científico

**Recibido:** 23/12/2025

**Aceptado:** 31/12/2025

**Publicado:** 28/02/2026

**Código Único AV:** e648

**Páginas:** 1(186-203)

**DOI:** <https://doi.org/10.5281/zenodo.18814655>

**Autor:**

**Ronald Verástegui Sánchez**

Contador Público

Doctor en Contabilidad

 <https://orcid.org/0000-0002-7226-3606>

**E-mail:** [revedi64@yahoo.com.ar](mailto:revedi64@yahoo.com.ar)

**Afiliación:** Universidad Tecnológica del Perú

**País:** República del Perú

**Isidora Concepción Zapata Periche**

Contador Publico

Doctor en Gestión y Ciencias de la Educación

 <https://orcid.org/0000-0001-9673-0435>

**E-mail:** [izapatap@unp.edu.pe](mailto:izapatap@unp.edu.pe)

**Afiliación:** Universidad Nacional de Piura

**País:** República del Perú

**Simón Zapata Periche**

Contador Publico

Doctor en Contabilidad y Finanzas

 <https://orcid.org/0000-0002-2214-1264>

**E-mail:** [speriche@unp.edu.pe](mailto:speriche@unp.edu.pe)

**Afiliación:** Universidad Nacional de Piura

**País:** República del Perú

**Rafael Jesús Herrera Torres**

Estudiante de Derecho

 <https://orcid.org/0000-0002-9105-8341>

**E-mail:** [rafaeljesusherrertorres1@gmail.com](mailto:rafaeljesusherrertorres1@gmail.com)

**Afiliación:** Universidad Cesar Vallejo

**País:** República del Perú

### Resumen

El presente artículo tuvo como objetivo analizar la integración de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de la deserción estudiantil en la educación superior, con el propósito de examinar cómo dichas dimensiones fueron incorporadas y valoradas en la literatura científica reciente. Para ello, se desarrolló una revisión sistemática siguiendo directrices metodológicas estandarizadas. La metodología empleada permitió sintetizar de manera estructurada los enfoques, modelos y estrategias de operacionalización de variables utilizados en la predicción de la deserción. Los resultados evidenciaron que los modelos que integraron de forma conjunta variables contextuales, conductuales y sociodemográficas tendieron a presentar un mayor valor explicativo y una mejor capacidad para identificar perfiles de riesgo, en comparación con aquellos basados exclusivamente en variables académicas. En conclusión, el estudio permitió reafirmar que la deserción estudiantil fue abordada de manera más efectiva desde una perspectiva multidimensional, destacándose la necesidad de desarrollar modelos predictivos más integrales, interpretables y contextualizados, capaces de orientar de forma pertinente las estrategias institucionales de prevención y retención en la educación superior.

### Palabras Clave

Deserción estudiantil, modelos predictivos, variables contextuales, variables conductuales, variables sociodemográficas.

### Abstract

Student dropout in higher education has been a complex and persistent problem affecting institutional efficiency, educational equity, and students' academic trajectories, prompting the development of predictive models aimed at its early detection. In this context, this article aimed to analyze the integration of contextual, behavioral, and sociodemographic variables in predictive models of student dropout in higher education, examining how these dimensions have been incorporated and valued in recent scientific literature. To this end, a systematic review was conducted following standardized methodological guidelines. The methodology employed allowed for a structured synthesis of the approaches, models, and strategies for operationalizing variables used in predicting dropout. The results showed that models integrating contextual, behavioral, and sociodemographic variables tended to have greater explanatory power and a better capacity to identify risk profiles compared to those based exclusively on academic variables. In conclusion, the study reaffirmed that student dropout was addressed more effectively from a multidimensional perspective, highlighting the need to develop more comprehensive, interpretable, and contextualized predictive models capable of appropriately guiding institutional prevention and retention strategies in higher education.

**Keywords** Student dropout, predictive models, contextual variables, behavioral variables, sociodemographic variables.

## Introducción

La deserción estudiantil (DE) en la educación superior se configura como un fenómeno de naturaleza compleja y multicausal, que se posiciona como uno de los retos de mayor relevancia para las instituciones educativas a escala global (Díaz-Landa et al., 2021). Esta problemática no solo incide en la calidad y continuidad de los procesos formativos, sino que también repercute de manera directa en los indicadores de eficiencia institucional, la sostenibilidad financiera y, de forma sustantiva, en las trayectorias académicas y profesionales de los estudiantes (Cárdenas et al., 2023). La magnitud del fenómeno se evidencia en las elevadas tasas de abandono registradas en diversos contextos geográficos y tipos de instituciones, lo que ha impulsado el diseño de estrategias de intervención sustentadas en la identificación temprana de variables asociadas al riesgo de deserción (Choque & Sosa-Jauregui, 2022).

En este escenario, la minería de datos educativos adquiere relevancia como un recurso metodológico con alto potencial para la identificación, el análisis y la eventual mitigación de la DE (Tavico, 2021). Los algoritmos de *machine learning* y las técnicas de *data mining* permiten procesar grandes volúmenes de información académica, conductual y sociodemográfica, con el fin de identificar patrones y características asociados al abandono de los estudios (Sánchez et

al., 2020). No obstante, la literatura científica actual evidencia una brecha significativa en la integración sistemática de variables contextuales, conductuales y sociodemográficas en los modelos predictivos desarrollados (Barahona-Anguisaca et al., 2024). Si bien existen estudios orientados a la predicción del rendimiento académico mediante técnicas de minería de datos, una proporción considerable de estos enfoques se concentra en variables académicas tradicionales, dejando de lado la complejidad multidimensional que caracteriza la DE (Cabral et al., 2024).

La incorporación limitada de estas dimensiones configura un vacío temático relevante, en la medida en que la DE deriva de la interacción de múltiples factores que exceden el ámbito estrictamente académico (Álvarez et al., 2022). En este sentido, las variables contextuales —como el entorno socioeconómico, la disponibilidad de recursos institucionales y el acceso a mecanismos de apoyo— se articulan con factores conductuales, tales como la motivación, la autorregulación del aprendizaje y los patrones de participación académica, así como con atributos sociodemográficos, entre ellos la edad, el género y el origen socioeconómico del estudiante (Caraballo, 2024).

La consideración fragmentada de estos componentes en los modelos predictivos restringe la capacidad institucional para diseñar intervenciones

personalizadas y contextualizadas que atiendan los factores subyacentes de la DE (Castro-Azuara et al., 2020). En consecuencia, resulta pertinente desarrollar una revisión sistemática orientada a integrar el conocimiento disponible, con el propósito de examinar cómo estas variables son incorporadas en los modelos de minería de datos destinados a la identificación temprana de la DE en la educación superior, reconociendo tanto los avances alcanzados como las limitaciones persistentes en este campo de estudio (Matas et al., 2020).

La literatura científica reciente documenta avances significativos en la aplicación de técnicas de minería de datos para la predicción de la DE en la educación superior; sin embargo, también revela limitaciones importantes en cuanto a la integración sistemática de variables contextuales, conductuales y sociodemográficas en los modelos desarrollados. Ayala et al., (2021) examinan modelos orientados a la estimación del riesgo académico en programas de computación mediante el uso de minería de datos aplicada al ámbito educativo, identificando atributos con elevada capacidad explicativa y proponiendo soluciones tecnológicas para la detección temprana del riesgo académico.

Aunque este estudio evidencia la factibilidad de emplear metodologías basadas en minería de datos para la identificación de estudiantes en situación de riesgo, su análisis se centra

predominantemente en variables académicas convencionales, relegando la complejidad multidimensional que caracteriza la DE.

De manera similar, Chacha et al., (2023) desarrollan un esquema predictivo de la DE sustentado en un modelo de regresión logística binaria múltiple, orientado a la identificación de situaciones potenciales de abandono en instituciones de educación superior en Ecuador. Si bien este trabajo constituye un aporte metodológico relevante al incorporar procedimientos estadísticos robustos, su alcance se limita al análisis de variables académicas y administrativas, sin integrar de forma articulada dimensiones contextuales, conductuales y sociodemográficas asociadas al proceso de decisión del estudiante respecto a la continuidad de sus estudios. Esta omisión restringe la capacidad explicativa del modelo y pone de manifiesto un vacío temático persistente en la literatura especializada.

Por su parte, Silva et al., (2020) analizan enfoques predictivos aplicados a la gestión de la deserción y la permanencia estudiantil, destacando el potencial de estos modelos para la formulación de políticas educativas y la detección temprana de estudiantes en riesgo de abandono. No obstante, aunque el estudio reconoce la influencia de múltiples factores en la DE, no profundiza en la forma específica en que variables contextuales, conductuales y sociodemográficas pueden

integrarse de manera sistemática en los modelos predictivos. Esta limitación refleja la brecha existente entre el reconocimiento teórico de la multidimensionalidad de la deserción y su operacionalización práctica en los modelos de minería de datos.

En conjunto, estos estudios evidencian que, pese al creciente interés en la aplicación de técnicas de minería de datos para la predicción de la DE, persiste una insuficiente integración de variables contextuales, conductuales y sociodemográficas en los modelos desarrollados. Esta carencia constituye un obstáculo significativo para el diseño de intervenciones personalizadas y contextualizadas que aborden las causas estructurales y psicosociales de la deserción, lo que subraya la necesidad de una revisión sistemática que sintetice el estado actual del conocimiento y proporcione directrices metodológicas para la incorporación efectiva de estas variables en futuras investigaciones.

La literatura científica sobre la predicción de la DE en educación superior mediante técnicas de minería de datos presenta, al menos, tres vacíos temáticos relevantes. En primer lugar, se observa una integración insuficiente de variables contextuales, conductuales y sociodemográficas en los modelos predictivos desarrollados. Gutiérrez et al., (2025) señalan que existe una escasa orientación hacia el análisis de factores socioconductuales con

capacidad predictiva, predominando modelos centrados en variables académicas tradicionales.

En segundo lugar, se identifica una carencia de estudios que analicen de manera sistemática el impacto de la inclusión de variables multidimensionales en la precisión y eficacia de los algoritmos predictivos. Rodríguez & Treviño (2024) reconocen la existencia de modelos que incorporan variables sociodemográficas, socioeconómicas e institucionales, pero advierten la ausencia de síntesis comparativas que evalúen su efectividad relativa frente a enfoques más restrictivos.

En tercer lugar, persiste una falta de directrices metodológicas claras y basadas en evidencia para la operacionalización e integración coherente de variables contextuales, conductuales y sociodemográficas. Si bien Kuz & Morales (2023) demuestran el potencial del *machine learning* para evaluar la permanencia estudiantil, no proporcionan orientaciones específicas sobre la selección, medición e integración sistemática de estas variables.

En este contexto, el presente artículo de revisión sistemática tiene como objetivo analizar la integración actual de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE en la educación superior, mediante la síntesis de la literatura científica publicada en los últimos diez años. Asimismo,

busca evaluar el impacto de estas variables en la precisión predictiva de los modelos y proponer directrices metodológicas orientadas a fortalecer futuras investigaciones en este campo.

### Metodología

La presente revisión sistemática fue conducida siguiendo las directrices del método PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Este enfoque metodológico garantizó la transparencia, la reproducibilidad y la calidad en la síntesis de la literatura científica.

La búsqueda en la base de datos Scopus se realizó mediante la aplicación de la siguiente fórmula booleana: (*"student dropout" OR "student desertion" OR "student attrition" OR "academic abandonment" OR "educational dropout"*) AND (*"data mining" OR "machine learning" OR "predictive model" OR "algorithm" OR "artificial intelligence"*) AND (*"higher education" OR "university" OR "tertiary education" OR "undergraduate" OR "postgraduate"*) AND (*"sociodemographic variable" OR "behavioral variable" OR "contextual variable" OR "socioeconomic factor" OR "demographic characteristic" OR "student characteristic"*))

La revisión sistemática fue guiada por tres preguntas de investigación específicas: a) ¿Cuál fue el estado actual de la integración de variables

contextuales, conductuales y sociodemográficas en los modelos predictivos de DE en la educación superior? b) ¿Cuál fue el impacto de la inclusión de variables contextuales, conductuales y sociodemográficas en la precisión y eficacia de los algoritmos de *data mining* y *machine learning* para la detección temprana de la DE, en comparación con modelos que utilizaron únicamente variables académicas? c) ¿Cuáles fueron las directrices metodológicas, mejores prácticas y recomendaciones que emergieron de la literatura para la operacionalización e integración efectiva de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE en la educación superior?

Los criterios de inclusión consideraron estudios publicados dentro del periodo temporal definido para la revisión, que abordaron la predicción o detección de la deserción estudiantil en el ámbito de la educación superior y que emplearon técnicas de *data mining* o *machine learning*. Asimismo, se incluyeron investigaciones que integraron al menos una variable de naturaleza contextual, conductual o sociodemográfica en sus modelos predictivos, que estuvieron disponibles en idioma inglés o español y que presentaron diseños metodológicos cuantitativos, cualitativos o mixtos. También se contemplaron artículos de investigación original, revisiones sistemáticas, metaanálisis y estudios de caso, siempre que aportaron evidencia

empírica o analítica relevante para el objetivo del estudio.

Por su parte, los criterios de exclusión comprendieron investigaciones centradas exclusivamente en la predicción del rendimiento académico sin considerar la DE, así como aquellas que utilizaron únicamente variables académicas sin integrar dimensiones contextuales, conductuales o sociodemográficas. Se excluyeron estudios desarrollados en contextos de educación primaria, secundaria o educación no formal, así como aquellos que no aplicaron técnicas de *data mining* o *machine learning*.

De igual modo, se descartaron publicaciones duplicadas o redundantes basadas en un mismo conjunto de datos, estudios con deficiencias metodológicas evidentes —entendidas como la falta de claridad en la metodología, en los criterios de selección de variables o en la validación de los modelos—, editoriales, comentarios, cartas al editor y trabajos sin acceso al texto completo. Finalmente, se excluyeron estudios publicados fuera del periodo establecido para la revisión o redactados en idiomas distintos del inglés y el español (Ver Figura 1).

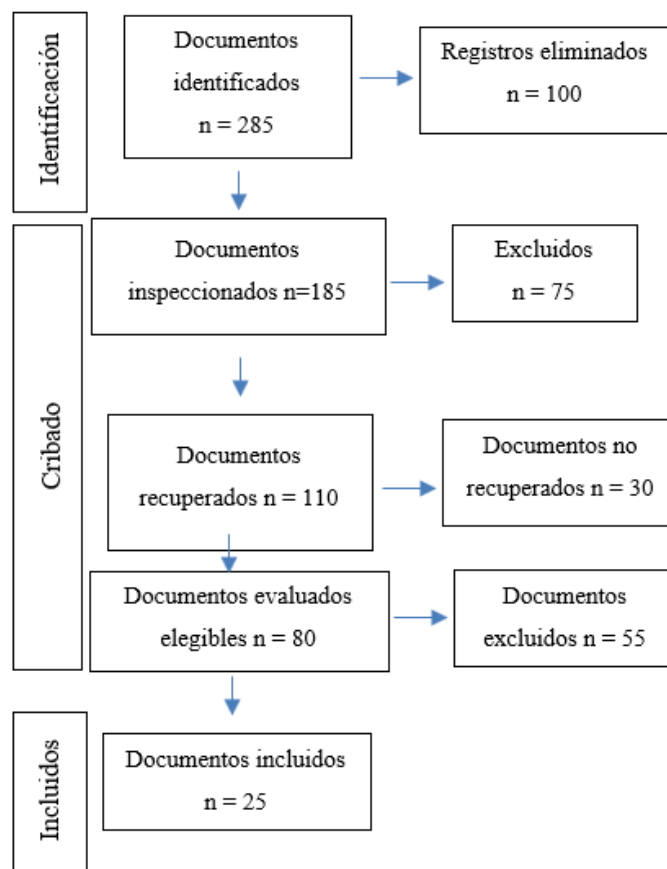


Figura 1. Identificación de estudios que utilizan el método prismático

## Resultados

Autor	Tipo de estudio / modelo	Variables contextuales	Variables conductuales y sociodemográficas	Principales hallazgos / aporte al objetivo
Aina et al., (2022)	Revisión sistemática socioeconómica	Institución, políticas educativas, condiciones económicas	Origen socioeconómico, antecedentes familiares, integración social	La deserción surge de la interacción entre factores individuales, institucionales y relacionales; la integración académica actúa como mediadora central
Contreras-Villalobos et al., (2024)	Métodos mixtos transformativos	Exclusión estructural, políticas públicas, territorio	Trayectorias educativas, conductas académicas, vulnerabilidad social	Evidencia un enfoque multinivel (macro-meso-micro) que integra variables contextuales y conductuales para explicar la exclusión y el abandono
Casanova et al., (2022)	Estudio empírico con intervención	Apoyo institucional, pandemia COVID-19	Motivación, adaptación académica, SES	Las variables conductuales (adaptación, emociones) y sociodemográficas inciden significativamente en el riesgo de deserción temprana
Mustofa et al., (2025)	ML híbrido + XAI (HLRNN)	Infraestructura educativa, políticas de retención	Rendimiento, salud mental, situación económica	Integra variables contextuales, conductuales y sociodemográficas con alta precisión y explicabilidad (SHAP/LIME)
Vaarma & Li (2024)	ML longitudinal (CAT, NN, LR)	Uso de LMS, estructura curricular	Actividad en LMS, créditos, demografía	La combinación de datos contextuales (LMS) y conductuales mejora la predicción, con efectos temporales relevantes
Szabó et al., (2024)	SEM y regresión	Clima escolar, apoyo familiar	<i>Engagement</i> (conductual, cognitivo, afectivo), educación parental	El <i>engagement</i> y la educación parental explican una proporción sustantiva del riesgo de abandono; destaca la complejidad multivariable
Wagner et al., (2024)	Estudio transversal con regresión logística	Condiciones del hogar, confinamiento, apoyo institucional	Inseguridad alimentaria, salud mental, género	La inseguridad alimentaria y el distrés mental incrementan significativamente la probabilidad de deserción universitaria
González-Morales et al., (2025)	Análisis de gran base de datos	Universidad, rama de estudios	Edad, sexo, empleo, becas, nacionalidad	Identifica perfiles de riesgo combinando variables sociodemográficas, académicas y laborales
Phan et al., (2023)	Framework DSS con datos textuales	Entorno institucional, feedback estudiantil	Opiniones textuales, rendimiento previo	Incorporar datos no estructurados (texto) mejora la predicción y segmentación del abandono
Coussement et al., (2020)	<i>Logit Leaf Model</i> (LLM)	Plataforma educativa, diseño del curso	<i>Engagement</i> cognitivo, académico y conductual	Segmenta alumnos y revela efectos diferenciales de variables conductuales y demográficas en la deserción
Vega-Rebolledo et al., (2025)	<i>Survival Analysis</i> + XAI	Programa académico, contexto institucional	Perfil de ingreso, créditos, GPA, edad	Muestra el carácter dinámico del abandono y la relevancia de variables de entrada y trayectoria con explicabilidad

**Tabla 1.** Integración de variables en modelos predictivos de deserción

Autor	Modelo / algoritmo	Tipo de variables integradas	Precisión / eficacia reportada	Aporte al objetivo de la investigación
Deleña et al., (2025)	XGBoost, GB, ANN, RF, SVM, LR, NB	Académicas + sociodemográficas	Exactitud 90.66 %, F1 = 90.72	Demostó que los modelos con variables sociodemográficas superan a los exclusivamente académicos en precisión predictiva
Kuzilek et al., (2021)	SVM, RF, NB, KNN	Conductuales (patrones de exámenes)	Incremento del F-measure $\approx$ 0.30	Evidenció que las variables conductuales mejoran la detección temprana del riesgo de abandono
Matz et al., (2023)	RF, GBM, SVM, LR	Académicas + sociodemográficas + conductuales ( <i>engagement</i> social)	AUC promedio 78 %, máximo 88 %	Confirmó que las variables conductuales y de experiencia universitaria aportan valor predictivo incremental
Mustofa et al., (2025)	Modelo híbrido HLRNN + SHAP/LIME	Académicas + socioeconómicas + conductuales	Exactitud 96 %	Mostró que la integración multivariable y la interpretabilidad aumentan la eficacia del modelo
Mallik & Gangopadhyay (2023)	Revisión de ML y DL en educación	Académicas, contextuales y sociodemográficas	Evidencia acumulada	Sustentó teóricamente que los modelos más eficaces integran dimensiones contextuales
Monteverde-Suárez et al., (2024)	ANN y <i>Naïve Bayes</i>	Académicas + sociodemográficas + entorno familiar	ANN con mejor sensibilidad y especificidad	Probó que la inclusión de variables sociodemográficas mejora la clasificación temprana de alumnos en riesgo
Phan et al., (2023)	LR + segmentación + doc2vec	Académicas + textuales ( <i>feedback</i> estudiantil)	Mejora del AUC y <i>top-decile lift</i>	Evidenció que las variables contextuales no estructuradas (texto) incrementan la precisión predictiva
Zheng et al., (2023)	ML evolutivo + DBN	Conductuales y contextuales ( <i>big data</i> educativo)	Mejora global del rendimiento del sistema	Sustentó el valor de integrar datos conductuales para sistemas inteligentes de predicción académica
Ragni et al., (2024)	Modelos multinivel con <i>propensity score</i>	Sociodemográficas + contextuales (modalidad híbrida)	Efectos diferenciales por programa	Evidenció que variables contextuales influyen indirectamente en desempeño y riesgo de deserción
Phan et al., (2023)	Framework DSS con segmentación	Académicas + sociodemográficas + textuales	Superioridad frente a <i>benchmarks</i>	Confirmó que la integración de variables contextuales complejas mejora la detección precoz de abandono

**Tabla 2.** Modelos predictivos de deserción

Autor	Enfoque metodológico / modelo predictivo	Variables contextuales, conductuales y sociodemográficas	Estrategia de operacionalización e integración	Directrices, mejores prácticas y recomendaciones
Maldonado et al., (2021)	Analítica predictiva orientada a decisiones	Participación en tutorías, rendimiento académico, contexto institucional	Integración de variables predictivas con métricas costo-beneficio	Alinear modelos predictivos con objetivos institucionales de retención
Rabelo & Zárate (2025)	Regresión logística, redes neuronales, árboles y <i>ensemble</i>	Factores personales, académicos, financieros y contextuales	Selección conceptual previa de variables	Adaptar los modelos al contexto institucional específico
Delogu et al., (2024)	<i>Random Forest</i> y <i>Gradient Boosting</i>	Ingreso familiar, tipo de escuela previa, desempeño del primer año	Uso de datos administrativos y <i>feature importance</i>	Priorizar desempeño temprano y variables socioeconómicas
Casanova et al., (2022)	Estudio aplicado psicoeducativo	Bienestar emocional, adaptación académica, contexto socioeconómico	Instrumentos de tamizaje integrados a datos académicos	Incorporar variables emocionales en sistemas preventivos
González et al., (2025)	Ecuaciones estructurales (SEM)	Variables académicas, psicológicas, sociales y económicas	Modelo mediado por agotamiento académico	Integrar variables multidominio y mediadores psicológicos
Aina et al., (2022)	Revisión sistemática socioeconómica	Factores familiares, institucionales y económicos	Marco teórico integrador	Evitar modelos unidimensionales
López-Angulo et al., (2023)	Modelo explicativo-predictivo (SEM)	Apoyo social, autoeficacia, satisfacción académica	Operacionalización de variables psicosociales	Priorizar variables conductuales modificables
Deleña et al., (2025)	Comparación de 10 algoritmos ML	Factores sociodemográficos y académicos históricos	CRISP-DM y validación cruzada	Preferir <i>ensemble</i> y <i>XGBoost</i> para alertas tempranas
Matz et al., (2023)	ML con datos macro, micro y meso-nivel	Socio-demografía, <i>engagement</i> , redes sociales	Fusión de datos institucionales y de interacción	Integrar <i>engagement</i> meso-nivel para mayor precisión
Szabó et al., (2024)	Regresión y SEM	Educación parental, <i>engagement</i> conductual, cognitivo y afectivo	Modelos multidimensionales de <i>engagement</i>	Tratar la deserción como proceso
Alshahrani (2025)	ML con SMOTE + XGBoost + XAI	Variables demográficas, educativas y del programa	Balanceo de clases y métricas múltiples	Abordar desbalance de clases y mejorar interpretabilidad
Mustofa et al., (2025)	Modelo híbrido HLRNN + SHAP/LIME	Factores sociales, económicos, académicos y conductuales	<i>Upsampling</i> , RFE-CV y XAI	Combinar precisión predictiva con explicabilidad
Delogu et al., (2024a)	ML predictivo con datos poblacionales	Rendimiento académico temprano y contexto familiar	Replicación con datos administrativos nacionales	Reafirma el valor de datos administrativos a gran escala
Delogu et al., (2024b)	Sistemas de alerta temprana con ML	Variables académicas iniciales y socioeconómicas	Integración longitudinal de cohortes	Diseñar políticas focalizadas basadas en riesgo temprano

**Tabla 3.** Integración de variables en modelos predictivos de DE

## Discusión de Resultados

Los resultados evidenciaron que la integración de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE en la educación superior se encontró en una fase de desarrollo desigual, caracterizada por avances metodológicos relevantes, pero también por persistentes vacíos de integración multidimensional. En coherencia con el objetivo del estudio, el análisis permitió identificar que los modelos que incorporaron de manera combinada estas categorías de variables tendieron a presentar un desempeño predictivo superior frente a aquellos enfoques centrados exclusivamente en variables académicas, lo que reforzó la comprensión de la DE como un fenómeno complejo y multicausal.

En relación con los resultados que mostraron una integración limitada de variables contextuales, conductuales y sociodemográficas en una parte significativa de los modelos revisados, los hallazgos convergieron con lo reportado por Ayala et al., (2021) y Chacha et al., (2023), quienes desarrollaron modelos predictivos con solidez estadística, pero con un énfasis predominante en variables académicas y administrativas. Esta convergencia permitió inferir que, pese al avance de las técnicas de minería de datos y aprendizaje automático, persistió una tendencia a priorizar indicadores de fácil acceso institucional, relegando variables psicosociales y contextuales cuya medición y

recolección demandaron mayores esfuerzos metodológicos.

De manera complementaria, los resultados de esta revisión se alinearon con estudios que evidenciaron los beneficios de una integración multivariable más amplia. Investigaciones como las de Aina et al., (2022) y Casanova et al., (2022) demostraron que los factores socioeconómicos, emocionales y de adaptación académica ejercieron una influencia significativa en el riesgo de DE, lo que fue consistente con los modelos analizados que incorporaron variables conductuales y sociodemográficas y reportaron mejoras sustantivas en su capacidad explicativa. Del mismo modo, los hallazgos coincidieron con Mustofa et al., (2025) y Matz et al., (2023), quienes evidenciaron que la combinación de variables académicas, conductuales y contextuales, junto con técnicas avanzadas de aprendizaje automático, incrementó la precisión predictiva y favoreció una segmentación más fina de los estudiantes en riesgo.

No obstante, se identificaron divergencias con estudios que, aun reconociendo teóricamente la naturaleza multidimensional de la DE, no lograron operacionalizar de manera sistemática dichas dimensiones en sus modelos. Trabajos como los de Silva et al., (2020) y Barahona-Anguisaca et al., (2024) destacaron la influencia de múltiples factores en el abandono, pero sin detallar estrategias claras para su integración en los algoritmos predictivos.

Estas diferencias se explicaron por limitaciones en la disponibilidad de datos, por enfoques metodológicos centrados en la gestión institucional más que en la predicción integral, o por la ausencia de marcos conceptuales que orientaran la selección y articulación coherente de variables contextuales y conductuales.

Asimismo, los resultados que evidenciaron una mejora en la precisión de los modelos al incorporar variables conductuales, como los patrones de participación en plataformas virtuales, los niveles de *engagement* académico y la autorregulación del aprendizaje, convergieron con los hallazgos de Kuzilek et al., (2021) y Vaarma & Li (2024). Esta coincidencia permitió sostener que los datos conductuales de carácter longitudinal aportaron información temprana y dinámica sobre el riesgo de abandono, reforzando la concepción de la DE como un proceso progresivo y no como un evento aislado.

A pesar de los aportes de esta revisión sistemática, se reconocieron limitaciones que condicionaron la interpretación de los resultados. En primer lugar, la revisión se circunscribió a estudios indexados en Scopus y publicados entre 2020 y 2025, lo que pudo haber excluido investigaciones relevantes disponibles en otras bases de datos o en literatura no indexada. En segundo lugar, se observó una elevada heterogeneidad metodológica entre los estudios incluidos, tanto en los algoritmos

empleados como en la definición y operacionalización de las variables, lo que dificultó la comparación directa de resultados y limitó la generalización de los hallazgos. En tercer lugar, muchos estudios priorizaron métricas de desempeño predictivo sin profundizar en la interpretabilidad de los modelos ni en su aplicabilidad práctica para el diseño de políticas institucionales de retención. Finalmente, se advirtió una representación limitada de contextos de países en desarrollo, lo que restringió la validez externa de los resultados en sistemas educativos con mayores niveles de desigualdad estructural.

A partir de los resultados obtenidos y de las limitaciones identificadas, futuras investigaciones deberán orientarse al desarrollo de marcos conceptuales integradores que guíen de manera explícita la selección, medición y articulación de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE. Asimismo, resultó pertinente promover diseños longitudinales y multinivel que permitieran capturar la evolución temporal del riesgo de abandono y la interacción entre factores individuales, institucionales y estructurales.

La incorporación de enfoques de inteligencia artificial explicable se perfiló como una línea prioritaria, al facilitar el equilibrio entre precisión predictiva e interpretabilidad, favoreciendo la adopción de estos modelos en la toma de decisiones

educativas. Finalmente, se recomendó ampliar la diversidad de contextos analizados y explorar el uso de fuentes de datos no estructuradas y variables psicosociales modificables, con el fin de fortalecer el desarrollo de sistemas de alerta temprana más sensibles, contextualizados y orientados a la prevención de la DE en la educación superior.

### Conclusiones

Los resultados de este estudio permitieron evidenciar que la literatura científica reciente coincidió en señalar que los modelos predictivos de DE en la educación superior alcanzaron mayores niveles de precisión, robustez y valor explicativo cuando integraron de manera conjunta variables contextuales, conductuales y sociodemográficas, en comparación con aquellos enfoques centrados exclusivamente en indicadores académicos tradicionales.

Los hallazgos más relevantes mostraron que las variables conductuales, tales como el *engagement* académico, los patrones de interacción en entornos virtuales y la autorregulación del aprendizaje, junto con variables sociodemográficas y contextuales —como el nivel socioeconómico, el apoyo institucional y las condiciones del entorno educativo— aportaron información sustantiva para la detección temprana del riesgo de abandono. En conjunto, estos resultados contribuyeron a reforzar la comprensión de la DE como un fenómeno multidimensional y dinámico, cuya predicción

requirió enfoques analíticos integradores y metodológicamente sólidos.

En relación con el objetivo de investigación planteado, el análisis realizado permitió concluir que la integración de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE en la educación superior se caracterizó por un desarrollo heterogéneo. Si bien se identificaron avances significativos en estudios recientes que adoptaron enfoques multivariados y técnicas avanzadas de aprendizaje automático, persistió una proporción considerable de investigaciones que continuaron priorizando variables académicas, principalmente por razones de disponibilidad de datos y simplicidad metodológica.

En este sentido, el estudio evidenció que la incorporación sistemática de variables multidimensionales no solo mejoró el desempeño predictivo, sino que también amplió la capacidad interpretativa de los modelos, favoreciendo la toma de decisiones informadas y el diseño de estrategias de intervención más contextualizadas y efectivas.

El presente trabajo se desarrolló como un artículo de revisión sistemática, lo que permitió sintetizar de manera rigurosa y transparente el estado del arte sobre la integración de variables contextuales, conductuales y sociodemográficas en los modelos predictivos de DE. Este enfoque metodológico permitió identificar patrones

comunes, divergencias conceptuales y metodológicas, así como vacíos de investigación en estudios empíricos previos, ofreciendo una visión estructurada y crítica de la evidencia disponible. Asimismo, la naturaleza sistemática de la revisión garantizó la trazabilidad de los resultados y su alineación con estándares internacionales de calidad científica.

Finalmente, las conclusiones del estudio permitieron reflexionar sobre la necesidad de avanzar hacia modelos predictivos más integrales, interpretables y sensibles a los contextos institucionales y socioculturales en los que se produce la DE. A partir de los resultados obtenidos, se identificó la pertinencia de orientar futuras investigaciones hacia el desarrollo de marcos conceptuales unificados para la operacionalización de variables multidimensionales, el uso de diseños longitudinales que capturen la evolución del riesgo de abandono y la incorporación de técnicas de inteligencia artificial explicable que faciliten la transferencia de los modelos a la práctica institucional.

Asimismo, se evidenció la necesidad de ampliar los estudios en contextos subrepresentados y de explorar nuevas fuentes de datos, con el fin de fortalecer la capacidad preventiva de los sistemas de alerta temprana y contribuir de manera efectiva a la mejora de la permanencia estudiantil en la educación superior.

## Referencias

- Aina, C., Baici, E., Casalone, G., & Pastore, F. (2022). The determinants of university dropout: A review of the socio-economic literature. *Socio-Economic Planning Sciences*, 79, 101102. Documento en línea. Disponible <https://doi.org/10.1016/j.seps.2021.101102>
- Alshahrani, A. (2025). SMOTE-optimized machine learning framework for predicting retention in workforce development training. *Computational Materials and Continua*, 85(2). Documento en línea. Disponible <https://doi.org/10.32604/cmc.2025.065211>
- Álvarez Quiñones, L. I., Lozano Moncada, C. A., & Bravo Montenegro, D. A. (2022). Metodología para el mantenimiento predictivo de transformadores de distribución basada en aprendizaje automático. *Ingeniería*, 27(3), e17742. Documento en línea. Disponible <https://doi.org/10.14483/23448393.17742>
- Ayala Franco, E., López Martínez, R. E., & Menéndez Domínguez, V. H. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *Revista de Educación a Distancia (RED)*, 21(66). Documento en línea. Disponible <https://doi.org/10.6018/red.463561>
- Barahona-Anguisaca, D., Vega-Calvas, P., Moyota-Paguay, A., & Porras-Ramírez, L. (2024). Análisis de modelos estadísticos para predecir el éxito académico en estudiantes universitarios. *MQR Investigar*, 8(2), 2951–2969. Documento en línea. Disponible <https://doi.org/10.56048/mqr20225.8.2.2024.2951-2969>
- Cabral Acosta, B. E., Cabral Acosta, C. L., & Cabral Acosta, K. R. (2024). Estrategias de retención estudiantil implementadas por la Facultad de Ciencias Contables, Administrativas y Económicas (2018–2022). *Ciencia Latina Revista Científica Multidisciplinar*, 8(4), 12440–12454. Documento en línea. Disponible [https://doi.org/10.37811/cl\\_rcm.v8i4.13446](https://doi.org/10.37811/cl_rcm.v8i4.13446)

- Caraballo Julio, N. R. (2024). El aprendizaje autónomo desde la perspectiva metacognitiva de Flavell. *Ciencia Latina Revista Científica Multidisciplinar*, 8(1), 10587–10603. Documento en línea. Disponible [https://doi.org/10.37811/cl\\_rcm.v8i1.10361](https://doi.org/10.37811/cl_rcm.v8i1.10361)
- Cárdenas Matute, J. M., Valle Franco, A., & Tapia Segarra, J. I. (2023). Factors that influence student desertion in the academic unit of social sciences of the Catholic University of Cuenca. *ConcienciaDigital*, 6(3), 30–48. Documento en línea. Disponible <https://doi.org/10.33262/concienciadigital.v6i3.2621>
- Casanova, J. R., Gomes, A., Moreira, M. A., & Almeida, L. S. (2022). Promoting success and persistence in pandemic times: An experience with first-year students. *Frontiers in Psychology*, 13, 815584. Documento en línea. Disponible <https://doi.org/10.3389/fpsyg.2022.815584>
- Castro-Azuara, M., Domínguez-Ángel, R., & Nava-Nava, R. (2020). Representaciones sobre prácticas letradas de estudiantes universitarios en situaciones vulnerables. *Magis. Revista Internacional de Investigación en Educación*, 12(25), 3–20. Documento en línea. Disponible <https://doi.org/10.11144/javeriana.m12-25.rple>
- Chacha, B., López, W., & Constante, M. (2023). Predictive model of student dropout based on logistic regression. *ESPOCH Congresses: The Ecuadorian Journal of STEAM*, 3(1), 630–656. Documento en línea. Disponible <https://doi.org/10.18502/epoch.v3i1.14477>
- Choque, V., & Sosa-Jauregui, V. (2022). Análisis del diseño curricular como factor de deserción académica utilizando minería de datos. *Yachay – Revista Científico Cultural*, 11(1), 551–555. Documento en línea. Disponible <https://doi.org/10.36881/yachay.v11i1.139>
- Contreras-Villalobos, T., Baleriola Escudero, E., & López Leiva, V. (2024). Digging and building: How transformative mixed-methods research contributes to explaining and responding to educational exclusion and school dropout. *Methods in Psychology*, 11, 100170. Documento en línea. Disponible <https://doi.org/10.1016/j.metip.2024.100170>
- Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020). Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, 135, 113325. Documento en línea. Disponible <https://doi.org/10.1016/j.dss.2020.113325>
- Deleña, R. D., Dia, N. J., Sacayan, R. R., Sieras, J. C., Khalid, S. A., Macatotong, A. H. T., & Gulam, S. B. (2025). Predicting student retention: A comparative study of machine learning approaches utilizing sociodemographic and academic factors. *Systems and Soft Computing*, 7, 200352. Documento en línea. Disponible <https://doi.org/10.1016/j.sasc.2025.200352>
- Delogu, M., Lagravinese, R., Paolini, D., & Resce, G. (2024). Predicting dropout from higher education: Evidence from Italy. *Economic Modelling*, 130, 106583. Documento en línea. Disponible <https://doi.org/10.1016/j.econmod.2023.106583>
- Díaz-Landa, B., Romero, R., & Rodríguez, W. (2021). Rendimiento académico de estudiantes en educación superior: Predicciones de factores influyentes a partir de árboles de decisión. *Telos*, 23(3), 616–639. Documento en línea. Disponible <https://doi.org/10.36390/telos233.08>
- González, B., Mendes, T. P., Pinto, R., Correia, S. V., Albuquerque, S., & Paulino, P. (2025). Predictors of higher education dropout intention in the post-pandemic era: The mediating role of academic exhaustion. *PLOS ONE*, 20(7), e0327643. Documento en línea. Disponible <https://doi.org/10.1371/journal.pone.0327643>
- González-Morales, M. O., López-Aguilar, D., Álvarez-Pérez, P. R., & Toledo-Delgado, P. A. (2025). Dropping out of higher education: Analysis of variables that characterize students who interrupt their studies. *Acta Psychologica*,

- 252, 104669. Documento en línea. Disponible <https://doi.org/10.1016/j.actpsy.2024.104669>
- Gutiérrez, A. P., Londoño, D. M. M., Landínez-Martínez, D., & Toro-Zuluaga, N. A. (2025). Las variables sociales y la conciencia metacognitiva de los jóvenes adultos colombianos. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 20(3), 1–32. Documento en línea. Disponible <https://doi.org/10.11600/rlcsnj.20.3.5379>
- Kuz, A., & Morales, R. (2023). Ciencia de datos educativos y aprendizaje automático: Un estudio de caso sobre la deserción estudiantil universitaria en México. *Education in the Knowledge Society (EKS)*, 24, e30080. Documento en línea. Disponible <https://doi.org/10.14201/eks.30080>
- Kuzilek, J., Zdrahal, Z., & Fuglik, V. (2021). Student success prediction using student exam behaviour. *Future Generation Computer Systems*, 125, 661–671. Documento en línea. Disponible <https://doi.org/10.1016/j.future.2021.07.009>
- López-Angulo, Y., Sáez-Delgado, F., Mella-Norambuena, J., Bernardo, A. B., & Díaz-Mujica, A. (2023). Predictive model of the dropout intention of Chilean university students. *Frontiers in Psychology*, 13, 893894. Documento en línea. Disponible <https://doi.org/10.3389/fpsyg.2022.893894>
- Maldonado, S., Miranda, J., Olaya, D., Vásquez, J., & Verbeke, W. (2021). Redefining profit metrics for boosting student retention in higher education. *Decision Support Systems*, 143, 113493. Documento en línea. Disponible <https://doi.org/10.1016/j.dss.2021.113493>
- Mallik, S., & Gangopadhyay, A. (2023). Proactive and reactive engagement of artificial intelligence methods for education: A review. *Frontiers in Artificial Intelligence*, 6, 1151391. Documento en línea. Disponible <https://doi.org/10.3389/frai.2023.1151391>
- Matas Terrón, A., Leiva Olivencia, J. J., & Núñez Rojas, N. (2020). El big data desde la perspectiva de estudiantes universitarios del norte del Perú. *Propósitos y Representaciones*, 8(1), e376. Documento en línea. Disponible <https://doi.org/10.20511/pyr2020.v8n1.376>
- Matz, S. C., Bukow, C. S., Peters, H., Deacons, C., Dinu, A., & Stachl, C. (2023). Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics. *Scientific Reports*, 13, 5705. Documento en línea. Disponible <https://doi.org/10.1038/s41598-023-32484-w>
- Monteverde-Suárez, D., González-Flores, P., Santos-Solórzano, R., García-Minjares, M., Zavala-Sierra, I., Luna de la Luz, V., & Sánchez-Mendiola, M. (2024). Predicting students' academic progress and related attributes in first-year medical students. *BMC Medical Education*, 24, 74. Documento en línea. Disponible <https://doi.org/10.1186/s12909-023-04918-6>
- Mustofa, S., Emon, Y. R., Mamun, S. B., Akhy, S. A., & Ahad, M. T. (2025). A novel AI-driven model for student dropout risk analysis with explainable AI insights. *Computers and Education: Artificial Intelligence*, 8, 100352. Documento en línea. Disponible <https://doi.org/10.1016/j.caeai.2024.100352>
- Phan, M., De Caigny, A., & Coussement, K. (2023). A decision support framework to incorporate textual data for early student dropout prediction in higher education. *Decision Support Systems*, 168, 113940. Documento en línea. Disponible <https://doi.org/10.1016/j.dss.2023.113940>
- Rabelo, A. M., & Zárate, L. E. (2025). A model for predicting dropout of higher education students. *Data Science and Management*, 8, 72–85. Documento en línea. Disponible <https://doi.org/10.1016/j.dsm.2024.07.001>
- Ragni, A., Ippolito, D., & Masci, C. (2024). Assessing the impact of hybrid teaching on students' academic performance. *Socio-Economic Planning Sciences*, 92, 101824.

- Documento en línea. Disponible <https://doi.org/10.1016/j.seps.2024.101824>
- Rodríguez, E. M., & Treviño, M. d. R. G. (2024). Causas de abandono y estrategias de permanencia universitaria en modalidad virtual: Revisión de literatura. *Ciencia Latina Revista Científica Multidisciplinar*, 8(5), 6217–6241. Documento en línea. Disponible [https://doi.org/10.37811/cl\\_rcm.v8i5.14036](https://doi.org/10.37811/cl_rcm.v8i5.14036)
- Sánchez Martínez, F., Barrera-Cámara, R., & Sáenz, A. (2020). Estudio evaluativo del sistema institucional de tutorías. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 10(20). Documento en línea. Disponible <https://doi.org/10.23913/ride.v10i20.609>
- Silva, F. C. da, Cabral, T. L. de O., & Pacheco, A. S. V. (2020). Dropout or permanence? Predictive models for higher education management. *Education Policy Analysis Archives*, 28, 149. Documento en línea. Disponible <https://doi.org/10.14507/epaa.28.5387>
- Szabó, L., Zsolnai, A., & Fehérvári, A. (2024). The relationship between student engagement and dropout risk in early adolescence. *International Journal of Educational Research Open*, 6, 100328. Documento en línea. Disponible <https://doi.org/10.1016/j.ijedro.2024.100328>
- Tavico Chamay, A. T. (2021). Factores que influyen en la deserción de la carrera de administración de empresas. *Revista Científica Internacional*, 4(1), 39–46. Documento en línea. Disponible <https://doi.org/10.46734/revcientifica.v4i1.45>
- Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society*, 76, 102474. Documento en línea. Disponible <https://doi.org/10.1016/j.techsoc.2024.102474>
- Vega-Rebolledo, I., Sánchez-García, A. J., Muñoz León, J. J., Ocharán-Hernández, J. O., & Cortés-Verdín, K. (2025). Applying survival analysis and explainable artificial intelligence to understand academic success. *Array*, 28, 100540. Documento en línea. Disponible <https://doi.org/10.1016/j.array.2025.100540>
- Wagner, F., Wagner, R. G., Makuapane, L. P., Masango, M., Kolanisi, U., & Gómez-Olivé, F. X. (2024). Mental distress, food insecurity and university student dropout during COVID-19. *Frontiers in Psychiatry*, 15, 1336538. Documento en línea. Disponible <https://doi.org/10.3389/fpsy.2024.1336538>
- Zheng, L., Wang, C., Chen, X., Song, Y., Meng, Z., & Zhang, R. (2023). Evolutionary machine learning builds smart education big data platforms. *Applied Soft Computing*, 136, 110114. Documento en línea. Disponible <https://doi.org/10.1016/j.asoc.2023.110114>