

# Using data mining to understand student attrition in universities: A systematic review

Ederick Gonzales, Bachelor<sup>1</sup>, Eduardo Colupú, Bachelor<sup>2</sup>

<sup>1,2</sup> Universidad Tecnológica del Perú, Peru, [U20202344@utp.edu.pe](mailto:U20202344@utp.edu.pe), [U20233512@utp.edu.pe](mailto:U20233512@utp.edu.pe)

*Abstract-- Student dropout in universities is a complex phenomenon influenced by academic, socioeconomic, demographic and institutional factors. This systematic review aims to analyze how data mining has been applied to predict and mitigate this problem, identifying relevant patterns in academic and demographic data. A total of 117 original articles found in databases such as Scopus, Dialnet and Ebsco were evaluated, of which 30 met the inclusion criteria. The most commonly used techniques included Random Forest, decision trees and XGBoost, standing out for their high accuracy in predicting attrition. The most effective predictive models identified at-risk students with an accuracy of over 90%, allowing personalized strategies to be designed and institutional resources to be optimized. Thus, data mining is an effective tool for anticipating and mitigating dropout, but its impact can be maximized by integrating it with qualitative approaches that consider psychosocial factors, which would favor more inclusive and effective interventions.*

*Keywords-- Data mining, Educational data mining, machine learning, university, student dropout, academic performance.*

# Uso de minería de datos para comprender la deserción estudiantil en universidades: Una revisión sistemática.

Ederick Gonzales, Bachelor<sup>1</sup>, Eduardo Colupú, Bachelor<sup>2</sup>

<sup>1,2</sup> Universidad Tecnológica del Perú, Peru, U20202344@utp.edu.pe, U20233512@utp.edu.pe

**Resumen**– La deserción estudiantil en universidades es un fenómeno complejo influenciado por factores académicos, socioeconómicos, demográficos e institucionales. Esta revisión sistemática tiene como objetivo analizar cómo la minería de datos ha sido aplicada para predecir y mitigar este problema, identificando patrones relevantes en datos académicos y demográficos. Fueron evaluados 117 artículos originales encontrados en bases de datos como Scopus, Dialnet y Ebsco fueron evaluados, de los cuales 30 cumplieron con los criterios de inclusión. las técnicas más utilizadas incluyeron Random Forest, árboles de decisión y XGBoost, destacando por su alta precisión en la predicción de deserción. Los modelos predictivos más eficaces identificaron estudiantes en riesgo con precisión superior al 90%, permitiendo diseñar estrategias personalizadas y optimizar recursos institucionales. Así, la minería de datos es una herramienta eficaz para anticipar y mitigar la deserción, pero su impacto puede maximizarse al integrarla con enfoques cualitativos que consideren factores psicosociales, lo que favorecería intervenciones más inclusivas y efectivas.

**Palabras clave**-- minería de datos, minería de datos educativos, aprendizaje automático, universidad, deserción de estudiantes, rendimiento académico.

## I. INTRODUCCIÓN

La deserción universitaria se refiere al cese de los estudios por parte del estudiante, ya sea de manera transitoria o definitiva. Esta problemática abarca una amplia gama de razones, que van desde factores individuales hasta ambientales. Se identifican al menos cuatro dimensiones que influyen en el proceso de deserción estudiantil: personal, institucional, laboral y familiar. Por lo tanto, la deserción universitaria no debe entenderse como un evento aislado con una única causa, sino como el resultado final de un proceso más largo y complejo en el que influyen diversas variables que llevan al estudiante a abandonar su carrera [33].

En el Perú, se tiende a pensar que la decisión de abandonar los estudios universitarios se debe únicamente a problemas económicos, donde los estudiantes se ven obligados a elegir entre estudiar y trabajar. Si bien este es el caso en la mayoría de las situaciones, existen otros factores que abarcan desde lo social hasta lo motivacional, los cuales están influenciados por el entorno local y regional [34]. Por ejemplo, un estudio realizado por la Superintendencia Nacional de Educación

Superior Universitaria (SUNEDU) entre 2012 y 2018 sobre la realidad universitaria en el Perú reveló que el porcentaje de deserción en la región selva fue del 24.6%, en la región sierra del 18.2% y en la región costa del 24%. Además, durante la pandemia del COVID-19, el 42.6% de los estudiantes universitarios provenientes de familias en situación de pobreza no extrema abandonaron sus estudios, mientras que el abandono en el nivel no pobre fue del 18.1% [34].

Estos datos evidencian la magnitud del problema y resaltan la necesidad de que las universidades no solo investiguen las causas de la deserción, sino que también se cuantifiquen los resultados obtenidos para formular estrategias efectivas que mitiguen este fenómeno. En este contexto, las técnicas de data mining se posicionan como herramientas esenciales por su capacidad para procesar y analizar grandes cantidades de datos generados en estos estudios.

La minería de datos es fundamental por su capacidad para descubrir patrones ocultos en grandes volúmenes de datos, lo que permite predecir la deserción estudiantil al identificar tempranamente a los estudiantes en riesgo de abandonar sus estudios. Esta técnica ha demostrado ser una herramienta eficaz para optimizar la toma de decisiones académicas y facilitar intervenciones oportunas [31]. Además, es clave para la predicción de la deserción estudiantil a través de los factores ya mencionados anteriormente, ya que el uso de modelos como árboles de decisión y regresiones logísticas ha logrado identificar con precisión a los estudiantes en riesgo, permitiendo la implementación de estrategias preventivas y reduciendo las tasas de abandono [32]. También permite predecir el rendimiento en cursos avanzados basados en prerrequisitos, lo cual es de gran utilidad para identificar a los estudiantes que podrían enfrentar dificultades en el futuro.

Este enfoque anticipado permite a las instituciones educativas detectar patrones que señalan posibles problemas de comprensión o desempeño, incluso antes de que los estudiantes inicien dichos cursos [32]. En vista de esta problemática, esta revisión de literatura tiene como objetivo explorar cómo la minería de datos puede transformar el enfoque educativo a través de la evaluación de casos similares en otros países,

mejorando así nuestra comprensión en esta área, sobretodo en su influencia tanto el rendimiento académico como en la retención de estudiantes.

## II. METODOLOGÍA

Durante el desarrollo de nuestra Revisión Sistemática de la Literatura (RSL), nuestro equipo se adhirió a la metodología PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), metodología la cual puede dividirse en 4 subprocesos: presentación de la lista de conceptos principales, acceso a la información vinculada a dichos conceptos, visualización de los documentos en su totalidad y visualización del informe final [35]. Iniciamos el proceso estructurando nuestra búsqueda empleando la estrategia PICO (Población, Intervención, Comparación, Resultados), lo cual nos permitió delimitar el alcance de nuestra investigación de manera precisa, definiendo nuestras preguntas de investigación y estableciendo tanto los criterios de inclusión como de exclusión [36].

**Tabla I – Definición de las preguntas PICO**

Tema de Investigación: <b>Uso de minería de datos para comprender la deserción estudiantil en universidades</b>	
Pregunta PICO: <b>¿Cómo la minería de datos ha influido en la predicción de deserción estudiantil en universidades?</b>	
<b>Acrónimo y componente</b>	<b>Subpreguntas</b>
<b>P</b> - Alumnos Universitarios	<b>PS1:</b> ¿Qué factores, según los estudios realizados, lleva a los alumnos universitarios a abandonar sus carreras?
<b>I</b> - Aplicación de data mining	<b>IS1:</b> ¿Cómo las técnicas de minería de datos han podido identificar con precisión a estudiantes que estén en riesgo de abandonar sus carreras universitarias?
<b>C</b> - Técnicas de minería de datos para el estudio del rendimiento académico	<b>CS1:</b> ¿Cuáles técnicas de minería de datos fueron más efectivas para el análisis y predicción del rendimiento académico en estudiantes universitarios?
<b>O</b> – Resultados de estudios basados en data mining en universidades	<b>OS1:</b> ¿Qué resultados se obtuvieron de estudios basados en minería de datos en universidades?

**Tabla II – Descripción de las palabras clave**

<b>Acrónimo y componente</b>	<b>Palabras clave</b>
<b>P</b> - Alumnos Universitarios	university AND students
<b>I</b> - Aplicación de data mining	application AND data AND mining
<b>C</b> - Técnicas de minería de datos para el estudio del rendimiento académico	techniques AND data AND mining AND academic AND performance
<b>O</b> – Resultados de estudios basados en data mining en universidades	data AND mining AND results AND universities

La búsqueda bibliográfica se realizó en bases de datos tales como Ebsco, Dialnet y Scopus, utilizando palabras clave y operadores booleanos. Aplicamos filtros temporales y eliminamos duplicados durante el cribado. Posteriormente, evaluamos críticamente los artículos seleccionados, considerando su metodología y relevancia. Finalmente, extrajimos y sintetizamos los datos relevantes, realizando un análisis cualitativo y cuantitativo para responder a nuestras preguntas de investigación.

Para las búsquedas en las bases de datos, se emplearon diferentes ecuaciones de búsqueda adaptadas a cada repositorio. Esta estrategia permitió optimizar el proceso de búsqueda y aumentar la eficiencia en la obtención de resultados relevantes. Las ecuaciones fueron las siguientes:

### **Scopus:**

((university OR students) AND (application OR data AND mining) AND (questionnaires OR measuring OR university OR comfort) AND (academic OR performance OR prediction) AND (dropout))

### **Dialnet:**

(Academic) AND (Students) AND (University dropout) AND (Data mining)

### **Ebsco:**

(Academic performance) AND (College students) AND (University dropout) AND (Data mining) AND (Results)

A continuación, se presenta una tabla que contiene los criterios de elegibilidad, los cuales son importantes para poder seleccionar la literatura que permita descartar estudios que no proporcionen información relevante para esta investigación.

**Tabla III – Descripción de los criterios de elegibilidad**

Nº	CRITERIOS DE INCLUSIÓN	Nº	CRITERIOS DE EXCLUSIÓN
CI1	Se incluirán investigaciones que utilicen conjuntos de datos relacionados con estudiantes, independientemente de la disciplina académica o del país de origen, siempre y cuando estén aplicados en contextos educativos.	CE1	Se excluirán investigaciones publicadas antes del año 2020.
CI2	Se considerarán estudios que empleen técnicas de minería de datos cuyo objetivo sea predecir el abandono estudiantil o evaluar el rendimiento académico.	CE2	Estudios que se centren en niveles educativos distintos al universitario (por ejemplo, estudios en educación primaria o secundaria).
CI3	Estudios que apliquen técnicas de data mining con el fin de ofrecer recomendaciones para la mejora de políticas educativas, la planificación institucional o el diseño de intervenciones que busquen reducir la deserción estudiantil.	CE3	Se excluirán artículos que no utilicen datos reales de estudiantes o que se centren exclusivamente en teorías sin realizar análisis empíricos de datos.

A partir de las búsquedas realizadas, se identificaron un conjunto de 117 artículos en total, distribuidos entre las bases de datos seleccionadas: de Scopus se obtuvieron 35 artículos, mientras que otros fueron recuperados de Dialnet (14 artículos) y Ebsco (98 artículos). Tras encontrar y eliminar 15 artículos duplicados, se procedió con el proceso de cribado inicial, donde se excluyeron aquellos que no se alineaban con la temática central de la revisión mediante la lectura de títulos y abstracts, se excluyeron 47 artículos en este proceso. Posteriormente, se seleccionó un grupo de artículos pertinentes, aunque algunos no estaban disponibles en texto completo, lo que redujo aún más la cantidad de estudios para evaluar su elegibilidad, siendo 3 artículos los excluidos. Durante esta fase, se aplicaron criterios de exclusión adicionales, eliminando aquellos cuya área de estudio donde se aplicaron técnicas de minería de datos no era la educativa, los que no se enfocaban en el análisis de datos de estudiantes, y aquellos publicados fuera del rango temporal

estipulado. Finalmente, un número reducido de artículos fue incluido en la revisión sistemática después de la lectura a texto completo. Estos fueron organizados y clasificados por autor, título y año de publicación, brindando una visión clara y ordenada de las contribuciones relevantes en el campo del uso de la minería de datos para la predicción del rendimiento académico, habiendo quedado finalmente con 14 artículos seleccionados de Scopus, 3 de Dialnet y 13 de Ebsco.

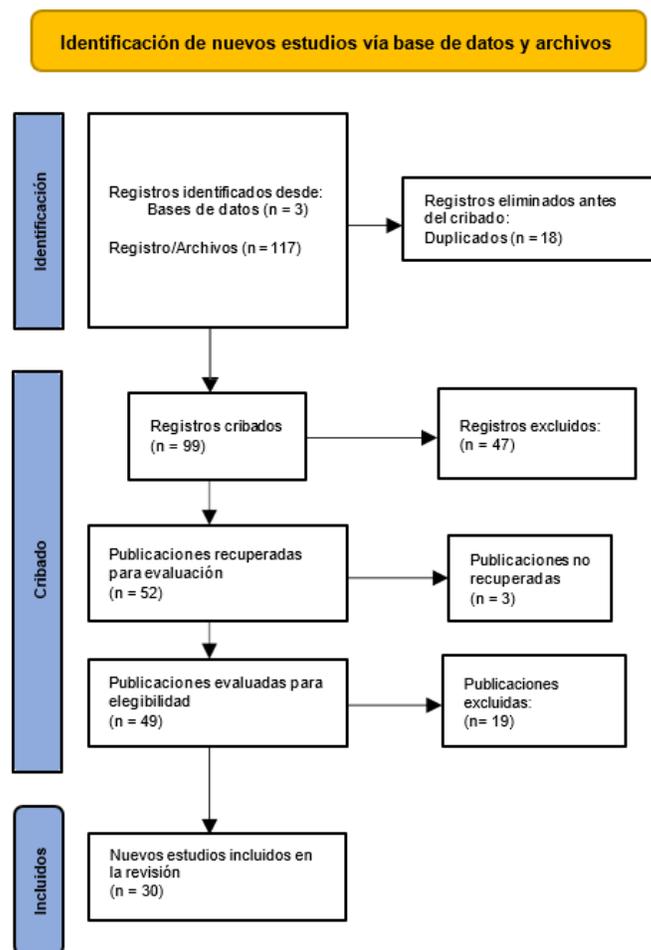


Fig. 1. Diagrama de flujo PRISMA

### III. RESULTADOS

Se presenta una síntesis general de los artículos revisados, organizados por autor, título del artículo y año de publicación. Esta estructuración proporciona una visión clara y organizada de los temas y aportes de cada artículo en el ámbito de la minería de datos y su empleo en el estudio de la deserción estudiantil en universidades.

En la figura 2, el gráfico representa la cantidad de artículos publicados anualmente entre 2020 y 2024. En 2020 se publicaron 7 artículos, en 2021 la cantidad disminuyó a 2, en

2022 subió a 8, y en 2023 se alcanzaron 10 publicaciones, mientras que en 2024 bajó nuevamente a 3.



Figura 2. Cantidad de artículos seleccionados según el año de publicación

En la figura 3, se muestra que el 93% de las publicaciones son "Artículos de Investigación", lo cual indica una fuerte inclinación hacia estudios completos que presentan hallazgos detallados. En comparación, solo el 7% corresponde a "Conference Papers", una proporción mucho menor.



Figura 3. Distribución de artículos seleccionados según el tipo de artículo

En la Figura 4, se observa la distribución de artículos publicados según su enfoque metodológico: cuantitativo y cualitativo. De los 30 artículos revisados, 29 corresponden a estudios cuantitativos, mientras que solo 1 es de tipo cualitativo.

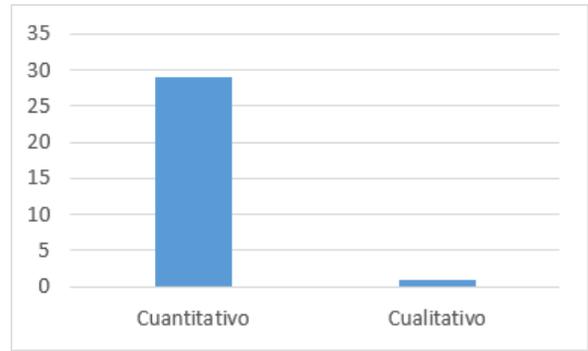


Figura 4. Distribución de artículos seleccionados según el tipo de estudio

En la Figura 5, se observa que, si bien la mayoría de los estudios están distribuidos en países de Latinoamérica, como Perú y Colombia con cinco artículos cada uno, también se evidencia una presencia significativa de estudios realizados en otras regiones, lo cual sugiere tanto una alta relevancia regional como un creciente interés internacional en abordar esta problemática.

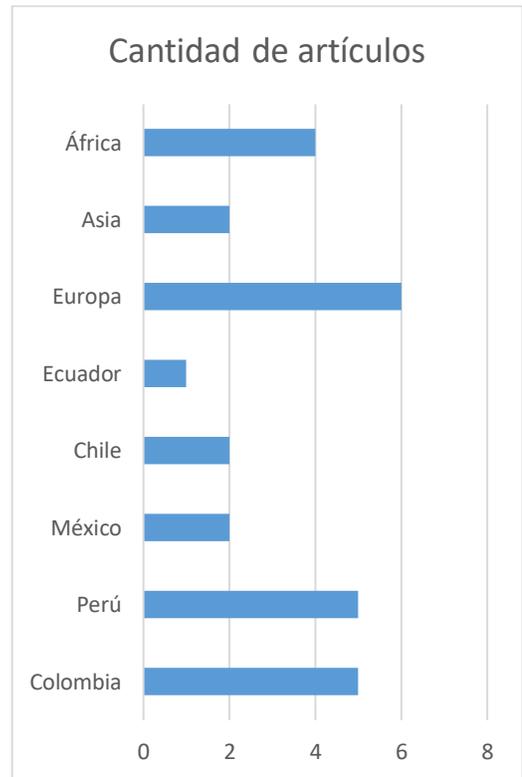


Fig. 5. Países de afiliación de los autores de los estudios incluidos en la revisión.

A. *¿Qué factores, según los estudios, llevan a los alumnos universitarios a abandonar sus carreras?*

La deserción estudiantil universitaria está determinada por múltiples factores interrelacionados. En el ámbito académico, el bajo rendimiento, reflejado en un promedio ponderado acumulado (GPA) reducido y una alta tasa de reprobación de asignaturas, constituye uno de los principales predictores de abandono [1, 4, 7, 9]. Esto se acentúa especialmente durante los primeros semestres [14, 25, 27] y cuando existe una inadecuada distribución de la carga académica [3, 30]. Las condiciones socioeconómicas desfavorables representan otro factor crítico, incluyendo limitaciones financieras [8, 15, 19], necesidad de trabajar durante los estudios [5, 27], y el estrato social de procedencia [15, 16]. Variables demográficas como la edad avanzada al ingreso [12, 14, 27], género [1, 12, 15] y origen geográfico [16, 27] también influyen significativamente. El contexto institucional juega un papel determinante a través de factores como la falta de asesoría y seguimiento personalizado [5, 8], insuficiente apoyo financiero [17, 28], y un ambiente universitario poco acogedor [8, 23]. Finalmente, aspectos personales como la baja autoeficacia [2], escasa motivación [10], frustración académica [10], limitada interacción con docentes [2, 11], y deficiente autogestión del aprendizaje [11, 29] completan el panorama multidimensional de este fenómeno.

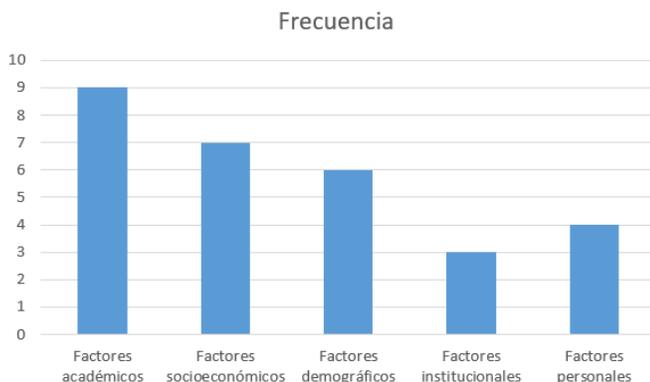


Figura. 6. Principales factores identificados en la literatura.

B. *¿Cómo las técnicas de minería de datos han podido identificar con precisión a estudiantes que estén en riesgo de abandonar sus carreras universitarias?*

La minería de datos ha permitido identificar con precisión a estudiantes en riesgo de abandono mediante diversas técnicas. Los árboles de decisión han demostrado ser particularmente efectivos, con algoritmos como C4.5 [5, 8] y J48 [10, 16] que analizan variables académicas y socioeconómicas, logrando precisiones de hasta 87.8% [16]. Los métodos de ensamble como Random Forest (RFC) han mejorado la segmentación por

programa, facultad y semestre [1, 4], mientras que técnicas como XGBoost han optimizado el análisis de datos específicos por programa académico [21, 28], alcanzando en algunos casos un área bajo la curva (AUC) de 96.1% [26]. Las redes neuronales, como el modelo CRRNN para analizar patrones de clics [11] y las DNN para evaluar calificaciones previas [24], han conseguido precisiones de hasta 89% en la identificación de estudiantes vulnerables. Técnicas de clustering jerárquico [12] y clasificación como Naïve Bayes y SVM [6, 17] han facilitado la segmentación basada en rendimiento y demografía. La integración de análisis de aprendizaje con sistemas de recomendación [13] y el uso de técnicas temporales como FCA [2] han mejorado la predicción en entornos virtuales. Metodologías como SMOTE han permitido balancear clases para obtener predicciones más precisas [20], mientras que herramientas como Watson Analytics [7] han facilitado la visualización de estos patrones. Los modelos más recientes combinan múltiples técnicas de clasificación [29] y se implementan a través de plataformas especializadas como H2O.ai [26] o herramientas web [25], logrando precisiones superiores al 90% en contextos específicos como ingeniería.

C. *¿Cuáles técnicas de minería de datos fueron más efectivas para el análisis y predicción del rendimiento académico en estudiantes universitarios?*

Las técnicas de minería de datos han demostrado diversa efectividad para predecir el rendimiento académico universitario. Los métodos de ensamble, particularmente Random Forest, destacan por su precisión superior, alcanzando un AUC de hasta 0.97 [1, 4, 7], especialmente en modelos segmentados por facultad o programa [4, 20]. Los árboles de decisión también muestran alta efectividad, con el algoritmo C4.5 logrando precisiones superiores al 90% [5, 8] y el J48 consiguiendo resultados similares [9, 10, 16, 27]. Algoritmos específicos como PART alcanzaron una precisión excepcional del 95.33% [14], mientras que GBM llegó a un AUC de 96.1% [26] y XGBoost a 87.75% [21, 28]. Las redes neuronales han ganado relevancia, con implementaciones DNN superando a métodos tradicionales con un 89% de precisión [24] y arquitecturas híbridas como CRRNN analizando efectivamente patrones de comportamiento a corto y largo plazo [11]. Las técnicas de clustering jerárquico [12] y K-means [6], especialmente cuando se combinan con clasificadores como SVM [25], han demostrado utilidad para segmentar perfiles de riesgo. Enfoques especializados como el Análisis de Conceptos Formales (FCA) con reglas temporales lograron F-scores del 85% en detección temprana de abandono [2]. La integración de técnicas en sistemas de recomendación híbridos [13] y la aplicación de métodos de balanceo como SMOTE [17, 20, 24] han mejorado significativamente los resultados. En contextos específicos, KNN alcanzó precisiones

excepcionalmente altas para clasificación por especialidad [18], mientras que técnicas exploratorias de visualización [23] y análisis estadísticos como regresión lineal y ANOVA [22] complementaron efectivamente los análisis predictivos en ciertos dominios.

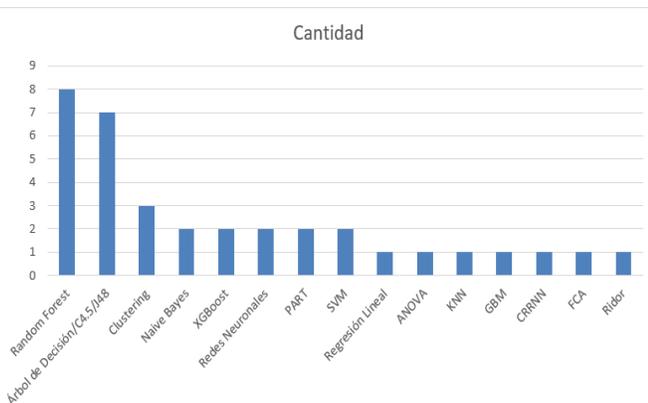


Figura. 7. Modelos más efectivos identificados en la totalidad de la literatura

#### D. ¿Qué resultados se obtuvieron de estudios basados en minería de datos en universidades?

La aplicación de minería de datos en contextos universitarios ha generado resultados significativos. El desarrollo de sistemas predictivos ha permitido la identificación temprana de estudiantes en riesgo [1, 4, 7], creando bases de conocimiento estratégicas [3] y perfiles de riesgo personalizados [10, 12] que facilitan intervenciones específicas. Las instituciones han implementado diversas herramientas prácticas, como aplicaciones móviles [5], sistemas de apoyo a la decisión para docentes [2, 9] y plataformas web [25], que han mejorado la capacidad de respuesta institucional. Los resultados incluyen mejoras en las estrategias de retención mediante intervenciones dirigidas a factores específicos como la pedagogía [8, 10], el ambiente estudiantil [8, 15] y el apoyo económico [9, 15]. En entornos de educación a distancia, estos sistemas han mostrado una aceptación positiva entre docentes (77%) [13] y han potenciado el análisis de interacciones en plataformas virtuales, alcanzando precisiones del 75% en la identificación de estudiantes en riesgo [11, 29]. Aplicaciones especializadas han optimizado procesos como la clasificación de estudiantes en especialidades adecuadas [18], la planificación de recursos para cursos críticos [26] y la adaptación de medidas preventivas para poblaciones específicas como estudiantes indígenas [16]. La implementación de estos modelos ha contribuido a una mejor asignación de recursos educativos [19, 21, 28], particularmente relevante durante la pandemia de COVID-19 [19], y ha permitido focalizar esfuerzos en etapas críticas como los primeros años académicos [27]. Los sistemas automatizados de detección [26] han demostrado ser adaptables a diferentes contextos institucionales [30], estableciendo un nuevo paradigma en la gestión educativa basada en evidencia.

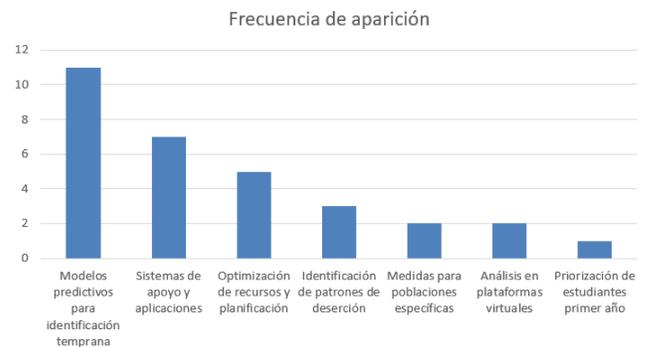


Figura.8. Medidas más comunes que se tomaron a partir del uso de modelos de minería de datos.

## IV. DISCUSIÓN

En esta revisión sistemática, se destacaron las técnicas de minería de datos como herramientas clave para predecir la deserción estudiantil con altas tasas de precisión, como es el caso de algoritmos como Random Forest y XGBoost, que lograron un AUC de hasta 0.97 [4, 7]. Estas técnicas permiten analizar múltiples factores, como variables demográficas, socioeconómicas, académicas e institucionales, ofreciendo predicciones robustas y adaptadas a distintos contextos educativos [8, 13]. Sin embargo, un aspecto crítico identificado en esta revisión es la necesidad de complementar estas técnicas con enfoques cualitativos que permitan un entendimiento más profundo de factores psicosociales, como motivación, estrés y percepción de apoyo institucional, los cuales son difíciles de capturar mediante análisis puramente cuantitativos [4]. La integración de estos enfoques podría optimizar la comprensión de la deserción estudiantil y permitir intervenciones más efectivas. Adicionalmente, varios estudios resaltaron la importancia de adaptar los modelos predictivos a las características específicas de cada población estudiantil, considerando contextos culturales y económicos que influyen directamente en las tasas de abandono [12, 16]. Esto refuerza la necesidad de diseñar estrategias personalizadas para mitigar la deserción, integrando no solo análisis predictivos, sino también medidas preventivas adaptadas a las necesidades locales. Finalmente, se observó una tendencia creciente en el uso de técnicas híbridas que combinan modelos de aprendizaje automático con métodos de análisis descriptivo, lo que ha permitido mejoras en la precisión y en la capacidad de generar recomendaciones específicas para las instituciones educativas [13, 20]. Estas técnicas no solo predicen el riesgo de deserción, sino que también ayudan a identificar estrategias para reducirlo, optimizando recursos y promoviendo ambientes de aprendizaje más inclusivos.

#### IV. CONCLUSIÓN

La deserción estudiantil universitaria emerge como un fenómeno multidimensional que trasciende la simple estadística para convertirse en un reflejo profundo de las complejas dinámicas sociales, económicas y existenciales que enfrentan los jóvenes en su tránsito hacia la realización profesional. Más allá de ser un indicador cuantitativo, representa una narrativa individual de esperanzas, limitaciones y resiliencia, donde cada abandono institucional constituye una historia personal de desafíos interrumpidos y potenciales truncados. Los factores que convergen en este proceso (académicos, socioeconómicos, personales e institucionales) revelan no solo barreras estructurales, sino también profundas reflexiones sobre el sentido de la educación superior en sociedades marcadas por desigualdades y transformaciones constantes.

Si bien técnicas como Random Forest, árboles de decisión y XGBoost han demostrado ser altamente eficaces para detectar patrones complejos en los datos, los resultados sugieren la necesidad de complementar estos enfoques cuantitativos con perspectivas cualitativas. La integración de factores psicosociales, como la motivación y la percepción del entorno educativo, permitiría desarrollar intervenciones más inclusivas que consideren la diversidad de contextos culturales y económicos de los estudiantes.

Finalmente, esta investigación evidencia que el abandono universitario no es un evento fortuito o aislado, sino el resultado de una compleja interacción entre aspiraciones individuales, condiciones materiales de existencia y capacidades institucionales para generar entornos de acompañamiento y desarrollo integral. Al adoptar un enfoque combinado de minería de datos y análisis cualitativo, se pueden crear soluciones más holísticas, optimizando recursos y fortaleciendo la retención y el éxito académico.

#### REFERENCES

- [1] C. Blundo, G. Fenza, G. Fuccio, V. Loia, y F. Orciuoli, "A time-driven FCA-based approach for identifying students' dropout in MOOCs", *Int J of Intelligent Sys.*, vol. 37, no 4, pp. 2683–2705, abr. 2022, doi: [10.1002/int.22414](https://doi.org/10.1002/int.22414).
- [2] P.-S. Deng y A. Chaudhury, "An Illustration of Using Adaptive Data Mining to Develop Strategic Knowledge Bases for Student Retention", 2023.
- [3] L. F. Castro Rojas, E. Espitia Peña, y E. Romero Cuero, "Análisis de características que influyen en la deserción estudiantil en el contexto de una universidad latinoamericana", *reveia*, vol. 20, no 40, dic. 2023, doi: [10.24050/reia.v20i40.1628](https://doi.org/10.24050/reia.v20i40.1628).
- [4] A. Jiménez-Macias, P. Moreno-Marcos, P. Muñoz-Merino, M. Ortiz-Rojas, y C. Kloos, "Analyzing feature importance for a predictive undergraduate student dropout model", *ComSIS*, vol. 20, no 1, pp. 175–194, 2023, doi: [10.2298/CSIS211110050J](https://doi.org/10.2298/CSIS211110050J).
- [5] E. F. Ruiz Ledesma, L. Chavarría Báez, y J. C. Veliz Martínez, "Aplicación móvil para estimar el porcentaje de deserción escolar usando minería de datos", *revEDU*, vol. 11, no 2, pp. 125–139, ago. 2023, doi: [10.37467/revedu.v11.4784](https://doi.org/10.37467/revedu.v11.4784).

- [6] H. Villarreal-Torres et al., "Classification model for student dropouts using machine learning: A case study", *EAI Endorsed Scal Inf Syst*, jun. 2023, doi: [10.4108/eetsis.vi.3455](https://doi.org/10.4108/eetsis.vi.3455).
- [7] S. M. Dol y P. M. Jawandhiya, "Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey", *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106071, jun. 2023, doi: [10.1016/j.engappai.2023.106071](https://doi.org/10.1016/j.engappai.2023.106071).
- [8] B. R. Pérez-Gutiérrez, "Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico", *revuin*, vol. 19, no 1, pp. 193–204, ene. 2020, doi: [10.18273/revuin.v19n1-2020018](https://doi.org/10.18273/revuin.v19n1-2020018).
- [9] V. Flores, S. Heras, y V. Julian, "Comparison of Predictive Models with Balanced Classes Using the SMOTE Method for the Forecast of Student Dropout in Higher Education", *Electronics*, vol. 11, no 3, p. 457, feb. 2022, doi: [10.3390/electronics11030457](https://doi.org/10.3390/electronics11030457).
- [10] M. Díaz-Choque et al., "Contributions of Data Mining to University Education, in the Context of the Covid-19 Pandemic: A Systematic Review of the Literature", *Int. J. Onl. Eng.*, vol. 19, no 12, pp. 16–33, ago. 2023, doi: [10.3991/ijoe.v19i12.40079](https://doi.org/10.3991/ijoe.v19i12.40079).
- [11] L. Blanquet, J. Grilo, P. Strecht, y A. Camanho, "Curbing Dropout: Predictive Analytics at the University of Porto", en *Atas da 23a Conferência da Associação Portuguesa de Sistemas de Informação, Associação Portuguesa de Sistemas de Informação, APSI*, 2023. doi: [10.18803/capsi.v23.143-159](https://doi.org/10.18803/capsi.v23.143-159).
- [12] A. B. Urbina-Nájera, J. C. Camino-Hampshire, y R. Cruz Barbosa, "Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa", *RELIEVE*, vol. 26, no 1, oct. 2020, doi: [10.7203/relieve.26.1.16061](https://doi.org/10.7203/relieve.26.1.16061).
- [13] A. Viloria et al., "Dropout-Permanence Analysis of University Students Using Data Mining", en *Intelligent Computing, Information and Control Systems*, vol. 1039, A.P. Pandian, K. Ntalianis, y R. Palanisamy, Eds., en *Advances in Intelligent Systems and Computing*, vol. 1039, Cham: Springer International Publishing, 2020, pp. 374–383. doi: [10.1007/978-3-030-30465-2\\_42](https://doi.org/10.1007/978-3-030-30465-2_42).
- [14] M. A. Prada et al., "Educational Data Mining for Tutoring Support in Higher Education: A Web-Based Tool Case Study in Engineering Degrees", *IEEE Access*, vol. 8, pp. 212818–212836, 2020, doi: [10.1109/ACCESS.2020.3040858](https://doi.org/10.1109/ACCESS.2020.3040858).
- [15] B. Assiri, M. Bashraheel, y A. Alsuri, "Enhanced Student Admission Procedures at Universities Using Data Mining and Machine Learning Techniques", *Applied Sciences*, vol. 14, no 3, p. 1109, ene. 2024, doi: [10.3390/app14031109](https://doi.org/10.3390/app14031109).
- [16] Vega, E. Sanz, P. De La Cruz, S. Moquillaza, y J. Pretell, "Intelligent System to Predict University Students Dropout", *Int. J. Onl. Eng.*, vol. 18, no 07, pp. 27–43, jun. 2022, doi: [10.3991/ijoe.v18i07.30195](https://doi.org/10.3991/ijoe.v18i07.30195).
- [17] K. M. Hassan, M. H. Khafagy, y M. Thabet, "Mining Educational Data to Analyze the Student's Performance in TOEFL iBT Reading, Listening and Writing Scores", *IJACSA*, vol. 13, no 7, 2022, doi: [10.14569/IJACSA.2022.0130741](https://doi.org/10.14569/IJACSA.2022.0130741).
- [18] L. Quiñones Huatangari, D. M. Jara, N. Alvarado, M. E. Milla, y O. A. Gamarra, "Modelo para la estimación de la deserción estudiantil Awajún y Wampis empleando minería de datos", *RECYT*, no 34, pp. 45–50, nov. 2020, doi: [10.36995/j.recyt.2020.34.006](https://doi.org/10.36995/j.recyt.2020.34.006).
- [19] J. Zárate-Valderrama, N. Bedregal-Alpaca, y V. Cornejo-Aparicio, "Modelos de clasificación para reconocer patrones de deserción en estudiantes universitarios", *Ingeniare. Rev. chil. ing.*, vol. 29, no 1, pp. 168–177, mar. 2021, doi: [10.4067/S0718-33052021000100168](https://doi.org/10.4067/S0718-33052021000100168).
- [20] O. D. Castrillón-Gómez, W. Sarache, y S. Ruiz-Herrera, "Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos", *Form. Univ.*, vol. 13, no 6, pp. 217–228, dic. 2020, doi: [10.4067/S0718-50062020000600217](https://doi.org/10.4067/S0718-50062020000600217).
- [21] J. H. Guanin-Fajardo, J. Guña-Moya, y J. Casillas, "Predicting Academic Success of College Students Using Machine Learning Techniques", *Data*, vol. 9, no 4, p. 60, abr. 2024, doi: [10.3390/data9040060](https://doi.org/10.3390/data9040060).
- [22] M. Jawthari y V. Stoffa, "Predicting At-Risk Students Using Weekly Activities and Assessments", *Int. J. Emerg. Technol. Learn.*, vol. 17, no 19, pp. 59–73, oct. 2022, doi: [10.3991/ijet.v17i19.131349](https://doi.org/10.3991/ijet.v17i19.131349).

- [23]I. A. Alhamad y H. P. Singh, "Predicting dropout at master level using educational data mining: A case of public health students in Saudi Arabia", *AI*, vol. 13, no 74, pp. 264–275, feb. 2024, doi: [10.34069/AI/2024.74.02.22](https://doi.org/10.34069/AI/2024.74.02.22).
- [24]X. Wang, B. Guo, y Y. Shen, "Predicting the At-Risk Online Students Based on the Click Data Distribution Characteristics", *Scientific Programming*, vol. 2022, pp. 1– 12, mar. 2022, doi: [10.1155/2022/9938260](https://doi.org/10.1155/2022/9938260).
- [25]A. Nabil, M. Seyam, y A. Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks", *IEEE Access*, vol. 9, pp. 140731–140746, 2021, doi: [10.1109/ACCESS.2021.3119596](https://doi.org/10.1109/ACCESS.2021.3119596).
- [26]M. Hinojosa, I. Derpich, M. Alfaro, D. Ruete, A. Caroca, y G. Gatica, "Procedimiento de agrupación de estudiantes según riesgo de abandono para mejorar la gestión estudiantil en educación superior", *Texto Livre*, vol. 15, p. e37275, mar. 2022, doi: [10.35699/1983-3652.2022.37275](https://doi.org/10.35699/1983-3652.2022.37275).
- [27][33]D. L. Morales Rodríguez, J. A. Caro Silva, y L. E. Contreras Bravo, "revisión sobre la predicción del rendimiento académico mediante métodos de ensamble", *ing. Solidar*, vol. 18, no 2, pp. 1–28, may 2022, doi: [10.16925/2357-6014.2022.02.01](https://doi.org/10.16925/2357-6014.2022.02.01).
- [28]M. Maphosa, W. Doorsamy, y B. S. Paul, "Student Performance Patterns in Engineering at the University of Johannesburg: An Exploratory Data Analysis", *IEEE Access*, vol. 11, pp. 48977–48987, 2023, doi: [10.1109/ACCESS.2023.3277225](https://doi.org/10.1109/ACCESS.2023.3277225).
- [29]A. I. Adekitan y O. Salau, "Toward an improved learning process: the relevance of ethnicity to data mining prediction of students' performance", *SN Appl. Sci.*, vol. 2, no 1, p. 8, ene. 2020, doi: [10.1007/s42452-019-1752-1](https://doi.org/10.1007/s42452-019-1752-1).
- [30]L. D. Andrade, C. M. M. D. Almeida, J. L. V. Barbosa, y S. J. Rigo, "Um modelo de Sistema de Recomendação integrado a Metodologias Ativas, MDE e Learning Analytics para a mitigação de evasão em EaD", *RELATEC*, vol. 22, no 2, pp. 185–205, jul. 2023, doi: [10.17398/1695-288X.22.2.185](https://doi.org/10.17398/1695-288X.22.2.185).
- [31]N. Roslan, J. M. Jamil, I. N. M. Shaharane, and S. J. S. Alawi, "Prediction of Student Dropout in Malaysian's Private Higher Education Institute using Data Mining Application," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 45, no. 2, pp. 168–176, Mar. 2025, doi: [10.37934/araset.45.2.168176](https://doi.org/10.37934/araset.45.2.168176).
- [32]Z. Alamgir, H. Akram, S. Karim, and A. Wali, "Enhancing Student Performance Prediction via Educational Data Mining on Academic Data," *Informatics in Education*, vol. 23, no. 1, pp. 1–24, 2024, doi: [10.15388/infedu.2024.04](https://doi.org/10.15388/infedu.2024.04).
- [33]J. A. Castro-Martínez y G. Machuca-Téllez, «La deserción universitaria en América Latina: una perspectiva ecológica», *Estud. pedagóg.*, vol. 49, n.o 2, pp. 87-108, 2023, doi: [10.4067/S0718-07052023000200087](https://doi.org/10.4067/S0718-07052023000200087).
- [34]J. I. Escalante López, C. J. Medina Valderrama, y A. Vásquez Muñoz, «La deserción universitaria: un problema no resuelto en el Perú», *AIAPÆC*, vol. 7, n.o 1, pp. 60-72, jun. 2023, doi: [10.26495/rch.v7i1.2421](https://doi.org/10.26495/rch.v7i1.2421).
- [35]E. Amigo, J. Gonzalo, V. Peinado, A. Penas, y F. Verdejo (2004) "PRISMA: un modelo interactivo de Síntesis de Información". ISSN 1135-5948, N° 33, 2004, págs. 9-16.
- [36]Nishikawa-Pacher, A. (2022). Research Questions with PICO: A Universal Mnemonic. Publications, 10(3). <https://doi.org/10.3390/publications10030021>