

Artículo de Investigación

Análisis Predictivo del Riesgo de Deserción: Un Enfoque Basado en Minería de Datos

Predictive Analysis of Attrition Risk: A Data Mining-Based Approach

Análise Preditiva do Risco de Evasão: Uma Abordagem Baseada em Mineração de Dados



Bibiana Díaz Sarmiento¹  , Maricela Morales Hernández¹  ,
Marisol Altamirano Cabrera¹  , Eva Rafael Pérez¹  ,
Claribel Benitez Quecha¹  , Aldo Jesús Chávez Cruz¹  

¹ Tecnológico Nacional de México, Instituto Tecnológico de Oaxaca, Departamento de Sistemas y Computación, México

Recibido: 2026-06-30 / **Aceptado:** 2026-07-30 / **Publicado:** 2026-08-15

RESUMEN

La deserción escolar en la Educación Superior constituye una problemática compleja que afecta significativamente el desarrollo de los estudiantes y la eficiencia operativa de las instituciones. Esta investigación, desarrollada en el Tecnológico Nacional de México-Instituto Tecnológico de Oaxaca, aborda dicho fenómeno en la carrera de Ingeniería en Sistemas Computacionales mediante un enfoque de análisis predictivo basado en minería de datos. Siguiendo el ciclo de vida de la metodología CRISP-DM, el estudio integró registros institucionales históricos con información recolectada a través de encuestas, evaluando de manera holística dimensiones académicas, socioeconómicas, psicológicas y de salud. Para transformar datos en conocimiento, se implementó una estrategia analítica híbrida: el algoritmo de agrupamiento K-Means para la segmentación del alumnado en niveles de riesgo (alto, medio y bajo), y el algoritmo Apriori para la extracción de reglas de asociación y patrones de comportamiento subyacentes. Los resultados permitieron caracterizar con precisión los factores multidimensionales que detonan la discontinuidad académica, lo que sirvió de base para el diseño de un sistema de alerta temprana. Esta propuesta proporciona a la gestión educativa una herramienta basada en evidencia empírica para fundamentar e implementar intervenciones preventivas, oportunas y personalizadas. Se concluye que la integración formal de la minería de datos en la Educación Superior, donde no solo fortalece la toma de decisiones estratégicas, sino que impulsa la retención estudiantil, reduce el rezago y optimiza la eficiencia terminal de Ingenieros en Sistemas Computacionales.

Palabras clave: deserción escolar, minería de datos, CRISP-DM

ABSTRACT

School dropout in Higher Education constitutes a complex issue that significantly impacts students' professional development and institutions' operational efficiency. Conducted at the Tecnológico Nacional de México-Instituto Tecnológico de Oaxaca, this research addresses this phenomenon within the Computer Systems Engineering program through a predictive analysis approach based on data mining. Following the lifecycle of the CRISP-DM methodology, the study integrated historical institutional records with data collected through surveys, holistically evaluating academic, socioeconomic, psychological, and health dimensions. For knowledge discovery, a hybrid analytical strategy was implemented: the K-Means clustering algorithm to segment students into risk levels (high, medium, and low), and the Apriori algorithm to extract association rules and underlying behavioral patterns. The results made it possible to accurately characterize the multidimensional factors that trigger academic discontinuity, serving as the foundation for designing an early warning system. This proposal provides educational management with a tool based on empirical evidence to support and implement preventive, timely, and personalized interventions. It is concluded that the formal integration of data

mining in Higher Education not only strengthens strategic decision-making, but also boosts student retention, reduces academic lag, and optimizes the graduation rate of Computer Systems Engineers.

Keywords: student dropout; data mining; CRISP-DM

RESUMO

A evasão escolar no Ensino Superior constitui uma problemática complexa que afeta significativamente o desenvolvimento profissional dos estudantes e a eficiência operacional das instituições. Esta pesquisa, desenvolvida no Tecnológico Nacional de México–Instituto Tecnológico de Oaxaca, aborda esse fenômeno no curso de Engenharia de Sistemas Computacionais por meio de uma abordagem de análise preditiva baseada em mineração de dados. Seguindo o ciclo de vida da metodologia CRISP-DM, o estudo integrou registros institucionais históricos com informações coletadas por meio de questionários, avaliando de forma holística dimensões acadêmicas, socioeconômicas, psicológicas e de saúde. Para a descoberta de conhecimento, foi implementada uma estratégia analítica híbrida: o algoritmo de agrupamento K-Means para a segmentação dos estudantes em níveis de risco (alto, médio e baixo), e o algoritmo Apriori para a extração de regras de associação e padrões de comportamento subjacentes. Os resultados permitiram caracterizar com precisão os fatores multidimensionais que desencadeiam a descontinuidade acadêmica, servindo de base para o projeto de um sistema de alerta precoce. Esta proposta fornece à gestão educacional uma ferramenta baseada em evidências empíricas para fundamentar e implementar intervenções preventivas, oportunas e personalizadas. Conclui-se que a integração formal da mineração de dados no Ensino Superior não apenas fortalece a tomada de decisões estratégicas, mas também impulsiona a retenção estudantil, reduz o atraso acadêmico e otimiza de forma substancial a taxa de diplomação em Engenharia de Sistemas Computacionais.

Palavras-chave: Evasão escolar; Mineração de dados; CRISP-DM

Forma sugerida de citar (APA):

Díaz Sarmiento, B., Morales Hernández, M., Altamirano Cabrera, M., Rafael Pérez, E., Benítez Quecha, C., & Chávez Cruz, A. J. (2026). Análisis Predictivo del Riesgo de Deserción: Un Enfoque Basado en Minería de Datos. *SAGA: Revista Científica Multidisciplinar*, 3(3), 282-290. <https://doi.org/10.63415/saga.v3i3.443>



Esta obra está bajo una licencia internacional
Creative Commons de Atribución No Comercial 4.0

INTRODUCCIÓN

La deserción universitaria en los programas de Ingeniería representa un desafío crítico en la Educación Superior, impactando negativamente los índices de eficiencia terminal y la optimización de recursos públicos (Martínez & Yeomans, 2023). En particular, la carrera de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México-Instituto Tecnológico de Oaxaca refleja esta problemática multicausal, condicionada por factores académicos, socioeconómicos, institucionales, motivacionales y de salud.

Frente a la necesidad de trascender los diagnósticos reactivos, la minería de datos educativa (EDM) emerge como un enfoque analítico robusto para descubrir patrones latentes de abandono. Esta investigación se presenta de forma estructurada bajo la metodología CRISP-DM, integrando datos

históricos institucionales y percepciones estudiantiles recopiladas. El análisis integra registros con encuestas aplicadas a estudiantes para evaluar diversas dimensiones académicas, socioeconómicas, psicológicas y de salud.

A través de la aplicación de técnicas avanzadas de minería de datos, específicamente los algoritmos de clustering (K-Means) y reglas de asociación (Apriori), se identifican perfiles de riesgo académico (alto, medio y bajo). El objetivo central es proponer un sistema de alerta temprana que oriente intervenciones preventivas personalizadas, optimizando la permanencia estudiantil y la toma de decisiones basada en evidencia dentro de la gestión universitaria.

METODOLOGÍA

Bases teóricas

El estudio de la trayectoria estudiantil en la Educación Superior requiere un marco

conceptual integrado que vincule la trayectoria con los factores de permanencia y abandono. En este sentido, la trayectoria escolar se concibe como un proceso continuo, completo y de excelencia, desarrollado sin interrupciones que comprometan el desempeño formativo (Secretaría de Educación Pública [SEP], 2024). El éxito de este trayecto se refleja en la eficiencia terminal, indicador que permite evaluar el grado en que los estudiantes concluyen su formación en los tiempos previstos (ANUIES, 2008).

La continuidad de esta trayectoria está condicionada a los procesos de permanencia y retención. Desde la perspectiva teórica de Tinto (1993), la persistencia académica depende directamente del grado de integración del estudiante, tanto en el plano académico (asociado al logro y desempeño) como en el social (vinculado al sentido de pertenencia). Cuando los lazos de integración se debilitan, se incrementa la vulnerabilidad del alumno.

En este contexto surgen desviaciones en el flujo escolar, tales como el rezago académico, el cual constituye un predictor directo del abandono al reflejar dificultades acumuladas desde el desempeño inicial (Asociación Nacional de Universidades e Instituciones de Educación Superior [ANUIES], 2008). La manifestación sostenida de este rezago sitúa a los alumnos en la categoría de estudiantes en riesgo, definidos como aquellos cuya probabilidad de concluir sus estudios en tiempo y forma es significativamente menor a la esperada para su cohorte. Esta condición suele estar multideterminada por factores como el bajo rendimiento, ausencias prolongadas, limitaciones socioeconómicas, afectaciones socioemocionales y barreras de adaptación institucional (Organisation for Economic Co-operation and Development [OECD], 2025).

Finalmente, la acumulación no atendida de estos factores de riesgo deriva en la deserción. Lejos de constituir un evento aislado, la deserción es el resultado final de un proceso paulatino de desintegración académica y social, en el cual el estudiante pierde

progresivamente el vínculo con la institución hasta consumir la decisión de abandonar sus estudios (Tinto, 1993).

Enfoque

En este contexto, esta investigación adoptó un enfoque cuantitativo, de alcance descriptivo-correlacional y de corte transversal. El proceso se estructuró bajo la metodología estándar para la minería de datos CRISP-DM (Cross-Industry Standard Process for Data Mining), organizada en seis fases iterativas: comprensión del dominio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

Población y Muestra

La población de estudio estuvo constituida por los estudiantes inscritos en la carrera de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México-Instituto Tecnológico de Oaxaca (ITO). Se empleó un muestreo no probabilístico por conveniencia, considerando dos fuentes principales de información:

- Registros académicos provenientes del sistema de control escolar institucional (Mindbox), abarcando variables de trayectoria académica (promedio acumulado, asignaturas reprobadas, avance de créditos y semestre cursado).
- *Cuestionario Estructurado*: Un instrumento elaborado con el propósito de capturar la percepción del estudiante en cinco dimensiones clave: socioeconómica, académica, institucional, motivacional y estado general de salud/bienestar.

Instrumento de recolección de datos

El cuestionario se diseñó utilizando reactivos cerrados con escalas Likert de 5 puntos (desde 1 = "Totalmente en desacuerdo" hasta 5 = "Totalmente de acuerdo") para medir la dimensión académica, institucional y motivacional. Para la salud física y mental se manejó 1=nunca y 5=siempre. Esta diferenciación responde a estándares psicométricos, donde la frecuencia es la forma correcta de evaluar bienestar/salud, mientras

que el acuerdo es la forma idónea de evaluar percepciones y actitudes.

Preparación y preprocesamiento de datos

Los datos brutos integrados de ambas fuentes fueron procesados mediante las siguientes etapas:

- a) *Limpieza e integración*: Se unificaron las bases de datos académicas y de encuestas utilizando la clave única de registro estudiantil como identificador primario. Se eliminaron los registros duplicados y aquellos que presentaban inconsistencias que pudieran afectar la calidad del análisis.
- b) *Tratamiento de valores faltantes y atípicos*: Se identificaron los registros con valores faltantes en variables críticas (p. ej., semestre o estatus académico), los cuales fueron eliminados. En las variables numéricas, como créditos aprobados y promedio, se evaluó su distribución y la presencia de valores atípicos antes de decidir su conservación o exclusión. En el caso del cuestionario, las respuestas incompletas fueron descartadas para reducir posibles sesgos analíticos. Este procedimiento sigue las recomendaciones de García et al. (2015), quienes señalan la importancia de tratar adecuadamente los valores faltantes durante el preprocesamiento de los datos.
- c) *Transformación y normalización*: Las variables créditos aprobados, créditos cursados y promedio certificado fueron normalizadas para evitar que las diferencias de escala influyeran de manera desproporcionada en el algoritmo de clustering. Esta práctica es recomendable en algoritmos basados en distancia, como K-means, debido a que variables con escalas mayores pueden ejercer una influencia superior en el cálculo de las distancias.
- d) *Codificación de variables categóricas*: Las variables categóricas fueron transformadas mediante técnicas de codificación adecuadas, mientras que las variables correspondientes a escalas tipo Likert del cuestionario se conservaron como ordinales, respetando su naturaleza y nivel

de medición. Este tratamiento busca preservar la información asociada al orden de las categorías y es consistente con las consideraciones metodológicas para el análisis multivariante de datos en ciencias sociales (Hair et al., 2019).

Técnicas de modelado

De acuerdo con los objetivos del proyecto y con la metodología CRISP-DM, en esta fase se aplican técnicas de minería de datos que permitan descubrir patrones, perfiles y relaciones relevantes en los datos académicos y complementarios de los estudiantes. Se seleccionó una combinación de técnicas no supervisadas, compuestas por:

- Clustering (K-Means), para la identificación de perfiles de riesgo académico
- Análisis de asociación, para detectar relaciones frecuentes entre variables académicas, socioeconómicas y motivacionales.
- Perfilado estadístico y visualización, como apoyo interpretativo de los resultados.

La elección de técnicas no supervisadas se justifica porque el fenómeno de la deserción no cuenta con una etiqueta predictiva directa en los registros y porque el interés principal es explorar estructuras latentes en los datos (Han et al., 2012).

Modelado mediante agrupamiento con el algoritmo K-Means

El algoritmo K-Means es ampliamente utilizado en la minería de datos educativos debido a su eficiencia computacional, interpretabilidad y escalabilidad en volúmenes masivos de datos numéricos (Jain, 2010; MacQueen, 1967).

En el contexto de esta investigación, K-Means permite agrupar a los estudiantes a partir de características académicas homogéneas. Esto facilita la identificación de perfiles asociados a niveles de riesgo bajo, medio y alto de deserción, mediante un enfoque no supervisado que prescinde de etiquetado o clasificación previa.

Para la construcción del modelo se seleccionaron cuatro variables académicas cuantitativas, extraídas del sistema institucional (Mindbox), identificadas en la literatura como predictores significativos del abandono escolar:

- Semestre cursado
- Créditos cursados
- Créditos aprobados
- Promedio académico certificado

Antes de ejecutar el algoritmo, las variables fueron estandarizadas mediante puntuación (Z-score), para prevenir sesgos asociados a las diferencias de escala intervariable. Este procedimiento es fundamental en algoritmos basados en distancia euclidiana para garantizar que todas las dimensiones aporten de manera equitativa al cálculo de proximidad (Han et al., 2012)

RESULTADOS

Se presentan y analizan los resultados obtenidos a partir de la aplicación de técnicas de minería de datos sobre la información institucional histórica (2015–2024) y los datos recolectados mediante el cuestionario estructurado aplicado a estudiantes del programa educativo de Ingeniería en Sistemas Computacionales del Instituto Tecnológico de Oaxaca.

Siguiendo la metodología CRISP-DM descrita en el capítulo anterior, los resultados se organizan conforme a las fases de modelado, evaluación y despliegue, integrando análisis cuantitativos, visualizaciones gráficas y una discusión interpretativa de los hallazgos, con el propósito de identificar patrones, perfiles de riesgo y factores asociados a la deserción escolar.

Resultados del análisis de clustering (K-Means)

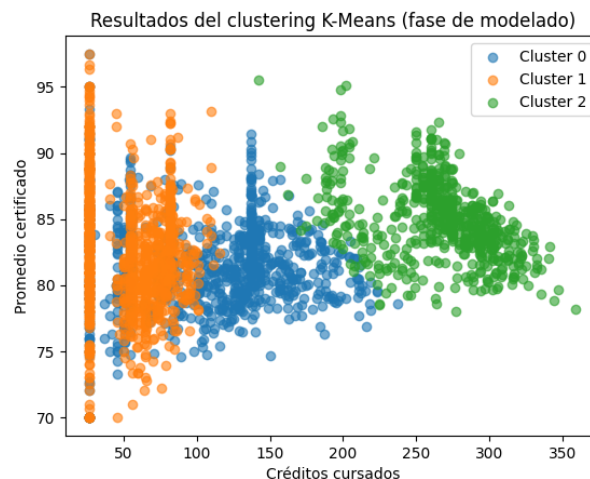
Identificación de clústeres académicos

La aplicación del algoritmo K-Means, utilizando como variables semestre, créditos aprobados, créditos cursados y promedio certificado, permitió segmentar a los estudiantes en tres clústeres claramente

diferenciados, los cuales fueron posteriormente interpretados como niveles de riesgo académico: riesgo alto, riesgo medio y riesgo bajo.

Figura 1

Resultados del clustering K-Means



Fuente: Elaboración propia

A partir del análisis de los centroides del modelo, se identificaron las siguientes características generales:

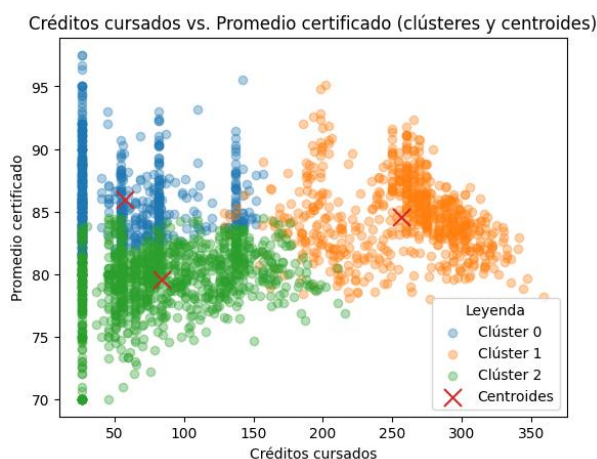
- *Clúster 0 de riesgo alto:* Estudiantes con menor número de créditos aprobados y cursados, ubicados en semestres relativamente bajos y con promedios académicos inferiores en comparación con los otros grupos. Este perfil refleja trayectorias académicas irregulares o con avances limitados, lo que incrementa la probabilidad de abandono.
- *Clúster 1 de riesgo medio:* Estudiantes con avances académicos moderados, que presentan un número intermedio de créditos cursados y aprobados, así como promedios aceptables pero inestables. Este grupo representa una población estratégica para intervenciones preventivas.
- *Clúster 2 de riesgo bajo:* Estudiantes con mayor avance académico, mayor acumulación de créditos aprobados y promedios relativamente estables. Aunque no se encuentran en riesgo inmediato, el análisis posterior permite identificar factores no académicos que podrían afectar su permanencia.

La distribución de estudiantes por nivel de riesgo, mostrada en la Figura 1, indica que una proporción significativa de la población se concentra en los niveles de riesgo alto y medio, lo cual confirma la relevancia institucional del problema de la deserción.

El Gráfico de dispersión: Créditos cursados vs. Promedio certificado, coloreado por clúster, evidencia una separación natural entre los grupos, lo que indica que las variables seleccionadas son adecuadas para la identificación de perfiles académicos diferenciados. La inclusión de los centroides permitió visualizar el comportamiento promedio de cada grupo, reforzando la interpretación de los clústeres.

Figura 2

Créditos cursados vs promedio certificado (clústeres y centroides)



Fuente: Elaboración propia

Los estudiantes pertenecientes a los clústeres con trayectorias académicas más estables presentan mayores niveles de acuerdo con la intención de continuar la carrera, mientras que los clústeres asociados a trayectorias irregulares o rezago académico concentran valores más bajos en la escala Likert. Esto confirma que el agrupamiento académico se relaciona directamente con variables motivacionales.

Este tipo de validación cruzada ha sido recomendado por Hernández (2020) como una práctica adecuada para evaluar modelos de minería educativa.

Resultados del análisis de la encuesta por clúster

Integración de datos institucionales y de percepción

Uno de los principales aportes del estudio fue la integración de los resultados del clustering con los datos obtenidos del cuestionario estructurado. Mediante el uso del número de control como identificador común, fue posible asociar a cada estudiante encuestado su respectivo nivel de riesgo académico.

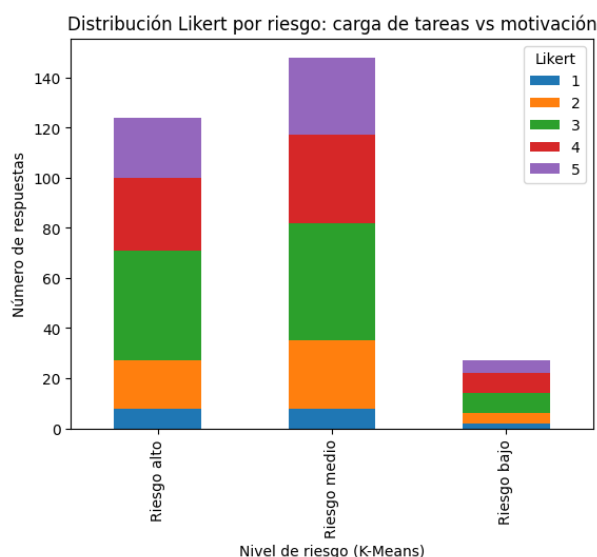
Esta integración permitió analizar cómo los factores académicos, motivacionales, institucionales y de salud se manifiestan de manera diferenciada entre los clústeres identificados.

Motivación y carga académica según nivel de riesgo

El análisis de la variable relacionada con la afirmación “La carga de tareas y proyectos afecta mi motivación para continuar” mostró diferencias relevantes entre los grupos.

El diagrama de caja por nivel de riesgo indica que los estudiantes clasificados en riesgo alto y medio presentan medianas más elevadas en esta variable, lo que sugiere una mayor percepción de sobrecarga académica y un impacto directo en su motivación. En contraste, el grupo de riesgo bajo muestra una distribución más concentrada en valores intermedios, reflejando una mejor adaptación a las exigencias académicas.

De manera complementaria, el gráfico de barras apiladas de la escala Likert evidencia que los niveles más altos de acuerdo (valores 4 y 5) se concentran principalmente en los clústeres de riesgo alto y medio, reforzando la relación entre carga académica percibida y riesgo de deserción.

Figura 3*Distribución Likert por riesgo**Fuente:* Elaboración propia

DISCUSIÓN

Los resultados obtenidos confirman que la deserción estudiantil no puede explicarse únicamente a partir de indicadores académicos tradicionales, sino que es un fenómeno multifactorial en el que interactúan variables académicas, motivacionales, institucionales y de bienestar.

El clustering permitió identificar perfiles objetivos basados en trayectoria escolar, mientras que la encuesta aportó información subjetiva clave, particularmente en aspectos como la motivación y el impacto emocional de la carga académica. Esta combinación de fuentes coincide con los planteamientos de Tinto (1993) y Cabrera et al. (1993), quienes conciben la permanencia estudiantil como un fenómeno multidimensional, influido tanto por la integración académica y social como por factores personales y ambientales. Desde esta perspectiva, la incorporación conjunta de indicadores académicos y variables psicosociales permite una comprensión más integral de los procesos asociados a la permanencia y al abandono estudiantil.

Asimismo, la existencia de estudiantes con riesgo académico bajo pero con señales de desmotivación o desgaste emocional sugiere que las estrategias institucionales de retención

no deben limitarse a estudiantes con bajo rendimiento, sino que deben contemplar acciones preventivas de acompañamiento integral.

Implicaciones institucionales

Los hallazgos de este estudio proporcionan insumos relevantes para la toma de decisiones institucionales, entre los que destacan:

- La posibilidad de identificar tempranamente estudiantes en riesgo mediante el uso de modelos de clustering aplicados a datos institucionales.
- El uso de la encuesta como herramienta complementaria para detectar factores no visibles en los sistemas escolares, como la motivación y la salud emocional.
- La segmentación de estrategias de intervención según el nivel de riesgo, priorizando acciones intensivas para los clústeres de riesgo alto y medio.

CONCLUSIONES

La presente investigación tuvo como objetivo aplicar técnicas de minería de datos para identificar patrones, variables relevantes y perfiles de riesgo asociados a la deserción académica en estudiantes del Programa Educativo de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México-Instituto Tecnológico de Oaxaca. A partir del análisis integrado de datos institucionales históricos (2015–2024) y de información recabada mediante un cuestionario estructurado aplicado a estudiantes activos, se logró cumplir de manera satisfactoria con los objetivos planteados.

Los resultados obtenidos permiten concluir que la deserción académica es un fenómeno multifactorial, en el que intervienen simultáneamente factores académicos, socioeconómicos, institucionales, motivacionales y de salud física y emocional. Ninguna variable, por sí sola, explica completamente el abandono escolar; sin embargo, la combinación de determinadas condiciones incrementa significativamente el riesgo.

La aplicación del algoritmo de clustering K-Means permitió identificar tres perfiles claramente diferenciados de estudiantes, los cuales fueron interpretados como clústeres de riesgo alto, medio y bajo. Estos perfiles no solo se distinguieron por variables académicas objetivas (créditos aprobados, promedio y avance curricular), sino que también mostraron diferencias relevantes en variables subjetivas provenientes de la encuesta, como la motivación, la percepción institucional y el bienestar emocional.

Asimismo, el análisis de reglas de asociación permitió identificar relaciones significativas entre variables, evidenciando que ciertos factores tienden a presentarse conjuntamente. Por ejemplo, se observó que estudiantes con alta carga académica y baja percepción de apoyo institucional presentan mayor probabilidad de reportar desmotivación y afectaciones emocionales, lo cual refuerza la idea de que la deserción es el resultado de interacciones complejas entre variables, más que de causas aisladas.

En conjunto, los resultados confirman la pertinencia del enfoque metodológico adoptado y demuestran que las técnicas de minería de datos constituyen una herramienta eficaz para el análisis de problemáticas educativas complejas, como la deserción universitaria.

Los hallazgos de esta investigación tienen implicaciones relevantes tanto a nivel académico como institucional. En el ámbito académico, los resultados evidencian la necesidad de considerar variables no tradicionales —como la salud emocional y la motivación— dentro de los modelos de seguimiento estudiantil, ya que estas influyen de manera significativa en la permanencia escolar.

A nivel institucional, los clústeres identificados pueden servir como base para el diseño de estrategias de intervención diferenciadas, permitiendo que los recursos disponibles se asignen de manera más eficiente. Por ejemplo, los estudiantes clasificados en el clúster de riesgo alto podrían beneficiarse de programas de tutoría intensiva,

apoyo psicológico y acompañamiento académico temprano, mientras que los de riesgo medio podrían ser atendidos mediante estrategias preventivas.

Además, el enfoque propuesto es replicable en otros programas educativos del Tecnológico Nacional de México, ya que se basa en datos institucionales comunes y en instrumentos fácilmente adaptables a otros contextos.

REFERENCIAS BIBLIOGRAFICAS

- Asociación Nacional de Universidades e Instituciones de Educación Superior [ANUIES]. (2008). Eficiencia en la educación superior: La necesidad de una nueva definición operacional. *Revista de la Educación Superior*, 37(146).
- Cabrera, A. F., Nora, A., & Castañeda, M. B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *The Journal of Higher Education*, 64(2), 123–139. <https://doi.org/10.1080/00221546.1993.11778419>
- García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. Springer.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Hernández, G. (2020). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Martínez-Líbano, J., & Yeomans-Cabrera, M.-M. (2023). Post-pandemic psychosocial variables affecting academic dropout in a sample of Chilean higher-education students. *Frontiers in Education*, 8, 1293259.

<https://doi.org/10.3389/feduc.2023.12932>
59

Organisation for Economic Co-operation and Development [OECD]. (2025). *Education at a glance 2025: OECD indicators*. OECD Publishing.

<https://doi.org/10.1787/1c0d9c79-en>

Secretaría de Educación Pública [SEP]. (2024). 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*, 14(28), e664.
<https://doi.org/10.23913/ride.v14i28.1923>

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press

AGRADECIMIENTO

Los autores hacen extensivo su agradecimiento al Tecnológico Nacional de México, Instituto Tecnológico de Oaxaca, Departamento de Sistemas y Computación por las facilidades otorgadas para realizar la investigación.

DECLARACIÓN DE CONFLICTO DE INTERESES

Los autores declaran no tener conflictos de intereses.



DERECHOS DE AUTOR

Díaz Sarmiento, B., Morales Hernández, M., Altamirano Cabrera, M., Rafael Pérez, E., Benitez Quecha, C., & Chávez Cruz, A. J. (2026)



Este es un artículo de acceso abierto distribuido bajo la licencia Creative Commons de Atribución No Comercial 4.0, que permite su uso sin restricciones, su distribución y reproducción por cualquier medio, siempre que no se haga con fines comerciales y el trabajo original sea fielmente citado.



El texto final, datos, expresiones, opiniones y apreciaciones contenidas en esta publicación es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la revista.