

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

Towards data-driven educational management: a platform for analyzing student dropout rates at CUJAE

Javier Molina Sánchez¹. Anaisa Hernández González². Gilda María Vega Cruz³

¹⁻³ Universidad Tecnológica de La Habana José Antonio Echeverría, Cuba

Correo electrónico:jmolina@ceis.cujae.edu.cu

ORCID <https://orcid.org/0009-0006-9378-1724>

Universidad Tecnológica de La Habana José Antonio Echeverría, Cuba

Correo electrónico:ahglez1970@gmail.com

ORCID <https://orcid.org/0000-0003-1169-301X>

Correo electrónico:gilda@tesla.cujae.edu.cu

ORCID <https://orcid.org/0000-0003-3363-7553>

Recibido: 28 de enero de 2026

Aceptado: 20 de marzo de 2026

Resumen

Este artículo aborda la transición hacia una gestión educativa basada en datos mediante el desarrollo e implementación de una plataforma digital destinada a optimizar el análisis y seguimiento de las bajas estudiantiles en la CUJAE. La solución debe reemplazar el sistema tradicional —basado en formularios físicos y hojas de cálculo, propenso a errores y de baja eficiencia— por una aplicación moderna que opere sin conexión y se sincronice posteriormente. La plataforma integra técnicas de inteligencia artificial y modelos no paramétricos para analizar las causas y patrones de deserción, permitiendo la generación de estadísticas en tiempo real. La metodología seguida comprendió: (1) diagnóstico del proceso manual, (2) selección de un stack tecnológico abierto (Vue.js, NestJS, SQLite, Prisma, Capacitor), (3) análisis y diseño de la solución, (4) implementación de la solución y (5) validación con datos reales. Los resultados validaron una reducción significativa de errores, una mejora en la eficiencia operativa y un incremento en la capacidad analítica institucional.

La solución no solo moderniza un proceso administrativo, sino que constituye un avance estratégico al proveer información oportuna y confiable para la toma de decisiones basada en evidencia y la elaboración de políticas educativas más efectivas.

Palabras clave: Transformación digital en la educación, deserción estudiantil, Plataforma digital, gestión educativa basada en datos.

Abstract

This work addresses the transition to data-driven educational management through the development and implementation of a digital platform designed to optimize the analysis and monitoring of student dropouts at CUJAE. The solution replaces the traditional system—based on physical forms and spreadsheets, prone to errors and inefficient—with a modern application that operates offline and synchronizes later. The platform integrates artificial intelligence techniques and non-parametric models to analyze the causes and patterns of dropout, enabling the generation of real-time statistics. The methodology followed comprised: (1) diagnosis of the manual process, (2) selection of an open technology stack (Vue.js, NestJS, SQLite, Prisma, Capacitor), (3) analysis and design of the solution, (4) implementation of the solution, and (5) validation with real data. The results validated a significant reduction in errors, an improvement in operational efficiency, and an increase in institutional analytical capacity. The solution not only modernizes an administrative process, but also constitutes a strategic advance by providing timely and reliable information for evidence-based decision-making and the development of more effective educational policies.

Keywords: Digital transformation in education, Student dropout, Digital platform, Data-driven educational management.

Licencia Creative Commons



Introducción

La transformación digital ha instaurado un paradigma educativo donde la automatización y el uso de datos son pilares para la mejora institucional. No obstante, en la Universidad Tecnológica de La Habana (CUJAE) persiste un método tradicional para el seguimiento de bajas, basado en papel y hojas de cálculo. Este enfoque manual es laborioso y presenta limitaciones pedagógicas: impide el análisis ágil de patrones de deserción, dificulta la retroalimentación oportuna para intervenciones y obstaculiza la generación de conocimiento institucional (Vega y *et. al.*, 2025). Así, la institución carece de la capacidad analítica necesaria para alinearse con los estándares actuales de gestión educativa basada en evidencia.

La deserción universitaria es un fenómeno crítico que afecta la eficiencia terminal en ingeniería y arquitectura en Cuba. Más que una cifra, implica una interrupción formativa que exige un análisis pedagógico profundo de sus causas e implicaciones sociales. Identificar sistemáticamente los motivos de abandono es un requisito pedagógico esencial para diseñar estrategias de retención efectivas y contextualizadas. Según González y *et al.* (2025), este enfoque no solo mitiga pérdidas cuantitativas, sino que eleva el rendimiento académico al atender trayectorias vulnerables y personalizar los apoyos institucionales. La literatura revela una carencia de soluciones para el seguimiento automatizado de exestudiantes (Del Bonifro et al., 2020; Niyogisubizo et al., 2022). Aunque existen sistemas de alerta temprana y aplicaciones de encuestas (Hernández et al., 2023; Borges et al., 2025; Quimiz-Moreira et al., 2025), son escasas las propuestas que integren recolección móvil, análisis avanzado y operación offline para entornos con conectividad intermitente. La brecha científica radica en la falta de herramientas flexibles que modelen dinámicamente las causas de deserción y sean interoperables.

La deserción estudiantil es un fenómeno complejo y multicausal (Del Bonifro y *et al.*, 2020; Borges y *et al.*, 2025) donde los métodos estadísticos descriptivos resultan limitados. Si bien cuantifican frecuencias, no logran identificar trayectorias no lineales, perfiles de riesgo emergentes ni interacciones dinámicas entre factores académicos, psicosociales e institucionales. Frente a ello, la minería de datos ofrece un avance metodológico clave: detecta relaciones ocultas, transforma datos en conocimiento accionable y permite pasar de una comprensión estática a un modelo sistémico y procesual que facilita intervenciones precisas y personalizadas (Dutt y *et al.*, 2017).

Los métodos no paramétricos aportan un valor pedagógico distintivo. Al no requerir supuestos distribucionales, trabajan con variables categóricas, ordinales o mixtas —comunes en educación— preservando la riqueza contextual. Son útiles para identificar patrones diferenciales entre subgrupos y asociaciones monotónicas sin linealidad, ofreciendo una comprensión más matizada de fenómenos complejos (Hollander y *et al.*, 2014). Las TIC facilitan la gestión eficiente de procesos educativos y administrativos. El análisis de datos y big data permite decisiones informadas, mejora de la calidad educativa, optimización de recursos y predicción de tendencias. Valencia y Almeida (2024) destacan que recopilar información sobre perfiles de ingreso, trayectorias y datos de egreso puede retroalimentar el sistema y generar mejoras. Gaftandzhieva *et al.* (2023) señalan que la gestión educativa debe basarse en datos mediante minería de datos educativos, análisis del aprendizaje e inteligencia empresarial. Esta integración transforma la gobernanza educativa, las metodologías de enseñanza, el liderazgo y la alfabetización digital (Kayumova y *et al.*, 2024).

En Cuba, la deserción ha sido abordada mediante estudios diagnósticos y descriptivos. En la CUJAE, un proyecto cuali-cuantitativo aplicó cuestionarios a estudiantes que causaron baja entre 2019 y 2022, priorizando su perspectiva fenomenológica para identificar factores de riesgo (Hernández *et al.*, 2023; Hernández y Ansola, 2024). Se analizaron variables como residencia, apoyo familiar, acceso a tecnología y motivación vocacional para correlacionarlas con la deserción. Por su parte, la Universidad de las Ciencias Informáticas realizó una investigación basada en caracterización psicopedagógica y rendimiento académico, clasificando las causas en voluntarias e involuntarias (Lázaro *et al.*, 2020). Ambos estudios usaron Excel para el procesamiento, pero sus hallazgos constituyen una base diagnóstica fundamental que subraya la necesidad de avanzar hacia modelos analítico-predictivos que conviertan la información en conocimiento pedagógico accionable.

Desde una perspectiva centrada en trayectorias educativas, identificar relaciones multicausales es clave para la tutoría proactiva y la gestión de la permanencia. Permite detectar constelaciones de factores de riesgo que, en interacción, elevan la probabilidad de abandono. Facilita una comprensión ecológica del fenómeno al desentrañar la dinámica entre causas proximales y distales, esencial para diseñar estrategias multidimensionales y personalizadas. Además, sienta las bases para un sistema de alerta temprana pedagógicamente informado que identifique oportunamente a estudiantes vulnerables y permita intervenciones preventivas sobre las causas raíz. La revisión bibliográfica confirma una brecha significativa que esta investigación busca cubrir: la ausencia de una plataforma

digital integral que automatice el seguimiento de la deserción, con capacidad offline, análisis multimétodo, integración de datos institucionales y representación flexible de la complejidad del fenómeno.

El objetivo general de esta investigación es diseñar e implementar una plataforma digital que optimice e; análisis y seguimiento de las bajas estudiantiles en la CUJAE. Esta solución busca superar las limitaciones del sistema manual vigente, estableciendo un sistema de información educativo que permita una recolección, gestión y análisis de datos más eficiente, seguro y pedagógicamente contextualizado. Su propósito último es generar conocimiento institucional accionable que sustente la mejora de las políticas de retención y el acompañamiento de las transiciones educativas. Este artículo describe la plataforma que se desarrolla con esta finalidad.

Materiales y métodos

La metodología de esta investigación se orientó al diseño, desarrollo e implementación de una plataforma digital integral para automatizar el seguimiento de estudiantes que causan baja en la Universidad Tecnológica de La Habana (CUJAE) (Rodríguez y Vega, 2020). El enfoque técnico adoptado integra el desarrollo web, una arquitectura modular, una base de datos relacional y capacidades de despliegue móvil, diseñado específicamente para operar en entornos con conectividad intermitente.

Este proceso implicó el diagnóstico del proceso manual, la selección de la tecnología para construir el producto informático, el análisis y diseño de la solución, su implementación y la validación empleando datos reales de las encuestas aplicadas a exestudiantes de los cursos 2019-2020, 2021 y 2022.

La solución se estructuró en tres componentes principales: (1) aplicación web de gestión dirigida a entrevistadores y administradores, permite gestionar cuestionarios, acceder a los datos y visualizar estadísticas, (2) aplicación móvil (APK) diseñada para la recolección de datos en campo, incluye funcionalidad de operación sin conexión y sincronización posterior y (3) base de datos local y centralizada que garantiza la persistencia, seguridad y flexibilidad necesarias para modelar la complejidad de los datos. El desarrollo técnico de la plataforma se sustentó en las siguientes tecnologías y componentes: el frontend se implementó con Vue.js, el backend se desarrolló utilizando NestJS y la para la base de datos se optó por SQLite, para la Sincronización y Empaquetado se utilizó Capacitor y el Componente de Inteligencia Artificial utilizó la plataforma KNIME Analytics.

Para el análisis de los datos recopilados, se seleccionó el algoritmo "Apriori" de minería de datos, basado en tres consideraciones fundamentales: la naturaleza categórica de las respuestas del cuestionario aplicado a los estudiantes; la capacidad del algoritmo para generar reglas de asociación intuitivas y fácilmente interpretables (en la forma {Antecedente} → {Consecuente}); y su idoneidad para identificar relaciones multivariadas, permitiendo detectar combinaciones complejas de factores que coexisten en los patrones de deserción.

La estadística no paramétrica es una rama de la inferencia estadística cuyo cuerpo de métodos se basa en modelos que no requieren que la función de distribución de probabilidad de la población esté completamente especificada. A diferencia de la estadística paramétrica, no depende de la estimación de parámetros fijos para realizar inferencias, sino que utiliza la estructura de los datos, como su orden o frecuencia, para contrastar hipótesis (Bautista et al., 2020). La aplicación de esta estadística en la investigación se justifica por:

- Naturaleza de las variables (escalas ordinales y cualitativas): El instrumento aplicado a los exestudiantes utiliza escalas ordinales (Buena, Regular, Mala) que carecen de una unidad de medida métrica constante. Por ello, se requiere un análisis basado en rangos y medianas en lugar de promedios.
- Tratamiento de muestras heterogéneas: El uso de técnicas no paramétricas ofrece una mayor resistencia a los valores atípicos presentes en los datos migrados mediante procesos de Extracción Transformación y Carga (ETL), permitiendo que el análisis de patrones sea más fiel a la realidad del comportamiento estudiantil.

Los algoritmos no paramétricos empleados fueron:

- Wilcoxon: Procedimiento estadístico diseñado para comparar dos conjuntos de datos relacionados o muestras pareadas. Evalúa si existen diferencias significativas entre las medianas de dos variables medidas en los mismos sujetos, sin asumir una distribución normal de los datos (Dao, 2022). Para los datos tomados como referencia, la hipótesis nula fue que no había diferencias entre la percepción de apoyo y la de adaptación de los estudiantes.
- Friedman: Compara tres o más mediciones relacionadas o muestras pareadas. A diferencia de los métodos que asumen parámetros fijos, este test utiliza la estructura de los datos para contrastar hipótesis basándose en rangos (Liu & Xu, 2022). En este caso el objetivo en su aplicación fue identificar científicamente cuáles dimensiones universitarias son las peor evaluadas por los estudiantes que causaron baja.

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

Kruskal-Wallis: Compara tres o más grupos independientes (Mckight & Najab, 2010). A diferencia de las pruebas que asumen parámetros poblacionales específicos, este método se basa en el análisis de rangos para determinar si las muestras provienen de la misma distribución. Al aplicarlo se buscó comparar científicamente la percepción de la "Calidad de Profesores" entre las diferentes carreras de la CUJAE.

Resultados

La plataforma digital desarrollada establece una relación de interoperabilidad con el Sistema de Gestión Universitaria (SIGENU- empleado en las universidades cubanas para gestionar el tránsito académico de un estudiante de pregrado desde la matrícula hasta su egreso o baja), del cual puede obtener automáticamente la información de los estudiantes que han causado baja. Adicionalmente, permite el registro manual de esta información de forma independiente. Para evaluar la solución, se implementó un proceso de Extracción, Transformación y Carga (ETL) que permitió migrar 419 registros desde archivos Excel hacia la base de datos relacional. La Tabla 1 compara la eficiencia entre el método tradicional y la plataforma digital.

Tabla 1 Comparación entre métodos de gestión. Fuente: elaboración propia

Método de gestión	Tiempo promedio por registro	Tasa de error	Capacidad de análisis
Manual (papel + Excel)	3.5 minutos	Alta	Limitada
Plataforma digital (ETL)	1.2 segundos	Nula	Avanzada (JSON + filtros)

Técnicamente, la solución implementa una arquitectura híbrida que integra una herramienta de análisis especializada (externa) con un sistema interno de almacenamiento, procesamiento y visualización de datos. Este diseño genera un flujo de información pedagógica integrado que permite actualizar periódicamente los hallazgos y visualizarlos en tiempo real a través de dashboards accesibles. La elección de esta arquitectura responde a un criterio de eficiencia pedagógico-tecnológica: permite validar rápidamente la utilidad y relevancia educativa del análisis de datos antes de comprometer recursos en el desarrollo de componentes nativos más complejos, asegurando así que la inversión tecnológica esté guiada por necesidades educativas demostradas y no por meras posibilidades técnicas.

En la implementación del algoritmo de minería de datos, se configuraron los umbrales de soporte mínimo en 1% y confianza mínima en 3%.

Esta parametrización se determinó considerando el volumen de datos disponible (aproximadamente 400-700 bajas por curso) y la necesidad de detectar asociaciones menos frecuentes, pero potencialmente significativas. Cada patrón identificado se complementa con sus métricas de frecuencia y fuerza asociativa. El análisis realizado a partir del procesamiento en KNIME sobre el conjunto de 419 cuestionarios completados, generó un total de 8 patrones. Entre los hallazgos más relevantes se encuentran:

- Reglas como: "*Si tipo_de_baja = 'Baja voluntaria', entonces causa incluye 'Salida del país'*" con un soporte del 48.8% y una confianza del 56.1%.
- "*Si tipo_de_baja = 'Baja por insuficiencia', entonces causa incluye 'No tenías suficiente preparación para enfrentar la carrera'*" con un soporte del 4.3% y una confianza del 33.3%.

En cuanto a los métodos no paramétricos, cada uno fue empleado con una finalidad diferente. Aunque los tres fueron implementados, aún no se incorporan a la solución informática pues están en proceso de evaluación los aportes que puedan suministrar.

La valoración de la aplicación de cada método permitió arribar a los resultados siguientes:

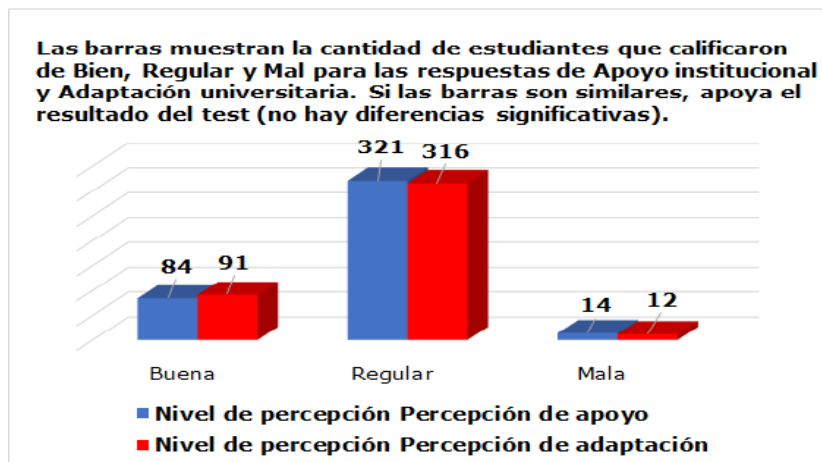
Wilcoxon

Variables claves para el análisis: percepción sobre la adaptación a la vida universitaria, percepción del apoyo de los profesores, apoyo del profesor guía y apoyo del jefe de año

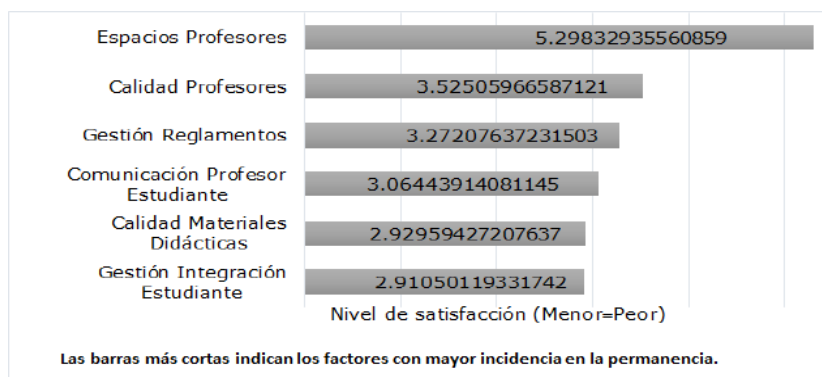
- Interpretación de resultados estadísticos: Resultados de la prueba de Wilcoxon para muestras pareadas: con 419 casos evaluados no se detectan diferencias significativas entre la percepción de apoyo institucional y la adaptación universitaria ($p = 0,3562$). La percepción de apoyo institucional y la de adaptación universitaria presentan patrones casi
- Interpretación contextual de resultados: Para la CUJAE, esto significa que la percepción de los estudiantes sobre el apoyo institucional está alineada con su percepción de adaptación. No se detecta una brecha significativa, lo cual indica una efectividad positiva del sistema de tutoría y acompañamiento actual. Sin embargo, este hallazgo no implica que el apoyo dénticos, reforzando que no existe diferencia significativa entre ambas variables haya sido suficiente para retener al estudiante. La deserción en estos 419 casos puede estar vinculada a factores externos (económicos o personales), confirmando que la adaptación no es el único determinante de la permanencia

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

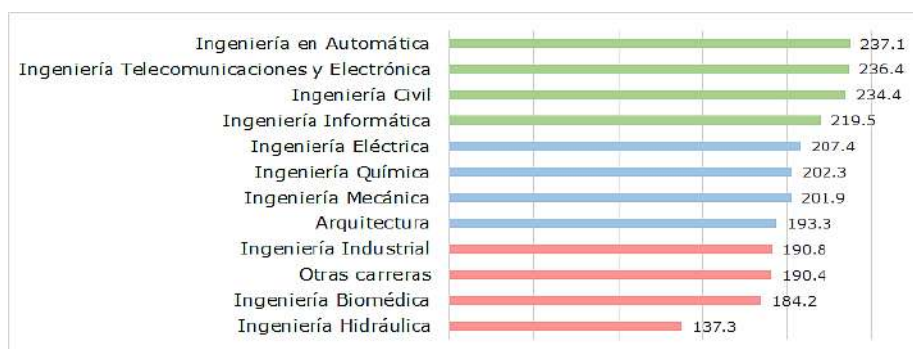
En la Figura 1 se grafican los resultados encontrados.



a)



b)



c)

Figura 1: Hallazgos de los métodos no paramétricos: a) Wilcoxon b) Friedman c) Kruskal-Wallis.

Friedman

- Variables claves para el análisis: calidad de profesores, calidad de materiales didácticos, comunicación profesor-estudiante, gestión de integración estudiante, gestión de reglamentos y espacios de actividades.

- Interpretación de resultados estadísticos: El test de Friedman revela diferencias significativas entre las dimensiones evaluadas ($Q = 496.713$, $p < 0.001$), confirmando que los estudiantes perciben desigualmente los factores institucionales.
- La Gestión de Integración y la Calidad de los Materiales Didácticos presentan los rangos medios más bajos, identificándose como los principales factores negativos en la retención estudiantil.
- Las comparaciones múltiples confirman que la Calidad de Profesores se valora significativamente mejor que la Gestión de Integración, los Materiales Didácticos y la Comunicación Profesor-Estudiante.
- Interpretación contextual de resultados: La Gestión de Integración y los Materiales Didácticos presentan los rangos medios más bajos, identificándose como los factores críticos que "empujan" al estudiante hacia la deserción. El análisis Post-Hoc confirma que la brecha entre la Calidad de Profesores y la Integración es significativa; el problema reside en el proceso de acogida y no en la competencia docente. Los Espacios de Actividades obtienen la mejor valoración, constituyéndose como una fortaleza institucional que puede utilizarse como estrategia de retención estudiantil.

Kruskal-Wallis

- Variables claves para el análisis: calidad de profesores y carrera.
- Interpretación de resultados estadísticos: Aunque las carreras de Ingeniería Hidráulica y Biomédica muestran los rangos más bajos, el contraste global no alcanza significación ($p = 0,087$), por lo que las diferencias inter-carreras en calidad docente no son estadísticamente concluyentes. Ingenierías Civil, Automática y Telecomunicaciones lideraron la evaluación positiva. Menos del 10% de estudiantes de Hidráulica califica la docencia como "Buena", frente al 44% de Automática, confirmando la brecha descriptiva aun sin significancia estadística.
- Interpretación contextual de resultados: A pesar de la falta de significancia global, los datos descriptivos muestran que Ingeniería Hidráulica (137) y Biomédica (186) presentan los rangos medios más bajos. Estas carreras requieren atención inmediata para evitar que la percepción negativa afecte el rendimiento académico. Aunque el test marca un $p = 0.087$, el ranking revela una brecha de 100 puntos de rango entre la carrera mejor evaluada y la peor. Esto sugiere que la desigualdad en la percepción es real y tangible para los estudiantes, aunque no haya superado el rigor del test de hipótesis.

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

- Las carreras de Ingeniería Civil, Automática y Telecomunicaciones y Electrónica (rangos 237-234) tiene las mejores percepciones. Sus métodos y prácticas docentes pueden ser analizados para evaluar su replicabilidad en otras carreras.

Discusión

El módulo de análisis de patrones de deserción desarrollado en esta investigación adopta un enfoque diagnóstico-pedagógico que prioriza la comprensión profunda del fenómeno sobre la predicción anticipada. Su propósito central es identificar relaciones estadísticamente significativas y patrones de asociación entre las variables categóricas recogidas mediante los cuestionarios aplicados a exestudiantes, empleando para ello técnicas de minería de datos exploratoria. Esta aproximación metodológica permite desvelar asociaciones ocultas y constelaciones causales que vinculan los motivos de baja declarados con los tipos de abandono observados, generando así un mapa etiológico estructurado y basado en evidencia. Este insumo resulta pedagógicamente valioso, ya que proporciona a los profesores, tutores y gestores una visión sistémica y fundamentada de los factores interrelacionados que configuran el fenómeno de la deserción, facilitando la transición desde percepciones anecdóticas hacia un diagnóstico institucional robusto que puede informar estrategias de intervención.

Los patrones identificados al aplicar el método A priori, permiten a los gestores del sistema identificar combinaciones de factores que se repiten con alta frecuencia, priorizar intervenciones institucionales basadas en evidencia y generar reportes dinámicos que resalten las asociaciones más fuertes.

En la Figuras 2 se aprecian los resultados obtenidos de procesar estos datos desde el Excel y empleando el método a priori.

Causas de las bajas voluntarias	Cantidad	%
Dificultades del transporte para llegar a la universidad	18	13%
No poseías una base sólida del nivel precedente	4	3%
No te gustó la carrera	34	25%
Problemas económicos	34	25%
Problemas personales	46	34%



a)

Rows: 8 | Columns: 6

Table Statistics

#	RowID	Support <small>Number (Float)</small>	Confidence <small>Number (Float)</small>	Lift <small>Number (Float)</small>	Consequent <small>String</small>	implies <small>String</small>	Items <small>Collection (Set)</small>
1	rule3	0.017	0.13	7.704	No tuviste tecnología disponibl	<--	[Baja por insuficiencia]
2	rule5	0.019	0.148	7.704	No tuviste apoyo de los profes:	<--	[Baja por insuficiencia]
3	rule7	0.036	0.278	7.704	Otras	<--	[Baja por insuficiencia]
4	rule9	0.043	0.333	7.704	No tenias suficiente preparació	<--	[Baja por insuficiencia]
5	rule11	0.058	0.066	1.149	Problemas económicos	<--	[Baja voluntaria]
6	rule13	0.065	0.075	1.149	No te gustó la carrera	<--	[Baja voluntaria]
7	rule15	0.079	0.091	1.149	Problemas personales	<--	[Baja voluntaria]
8	rule17	0.488	0.561	1.149	Salida del país	<--	[Baja voluntaria]

b)

Figura 2: Procesamiento en Excel (a) y empleando la plataforma (b).

Los hallazgos de esta investigación validan la premisa central del estudio: el diseño e implementación de una plataforma digital integral, desarrollada con un stack tecnológico contemporáneo y contextualmente adaptado a la infraestructura y los procesos de la CUJAE, produce una mejora cuantificable y sustantiva en la eficiencia operativa del seguimiento de la deserción y, de manera crucial, en la profundidad y confiabilidad del análisis causal. La informatización del flujo de trabajo, la migración estructurada de datos históricos y la validación exhaustiva del sistema a través de pruebas funcionales y de usuario, demuestran que la solución desarrollada responde de manera efectiva y robusta a las necesidades operativas y analíticas específicamente diagnosticadas en la institución.

Un análisis comparativo entre la nueva plataforma y el método tradicional —basado en formularios físicos, comunicación no estructurada y procesamiento manual en hojas de cálculo— revela avances transformadores en indicadores clave. La eficiencia del procesamiento experimentó una optimización de orden de magnitud: el tiempo medio requerido por registro se redujo de aproximadamente 3.5 minutos a 1.2 segundos, lo que representa una mejora superior al 98% en la velocidad de gestión de la información. Paralelamente, se logró eliminar la tasa de error en la transcripción y consolidación de datos, erradicando las inconsistencias, duplicaciones y pérdidas de información que eran endémicas en el proceso anterior y que comprometían la integridad de cualquier análisis posterior. Cabe destacar que, a diferencia de intentos parciales de digitalización documentados en el contexto, esta solución constituye una digitalización integral del flujo end-to-end, abarcando desde la captura inicial de datos y la gestión de permisos por roles, hasta el almacenamiento seguro,

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

el análisis estadístico avanzado y la visualización interactiva de resultados, estableciendo así un nuevo estándar de gestión de la información para el fenómeno de la deserción.

En cuanto a los métodos no paramétricos implementados, los resultados revelan que pueden ofrecer información relevante para asociada a los datos estudiados que pueden ser referente para estrategias específicas. El test de Kruskal Wallis convirtió respuesta de encuesta en un ranking claro de carreras por percepción docente. El test de Friedman permitió transformar datos de encuesta en decisiones gerenciales asociadas con la necesidad de reformar los procesos de integración de los estudiantes y disponibilidad de materiales didácticos. El test de Wilcoxon evidenció ausencia de diferencias significativas entre dos aspectos claves del acompañamiento estudiantil: la percepción de apoyo y la percepción de adaptación.

El proceso de implementación y validación permitió identificar limitaciones técnicas y operativas relevantes para la sostenibilidad y escalabilidad de la solución. En el ámbito técnico, se identificó una dependencia crítica del sistema operativo, ya que la plataforma requiere dispositivos móviles con Android 8 (API 26) o superior, lo que podría generar una brecha de acceso tecnológico y restringir su cobertura en contextos institucionales con parques de dispositivos heterogéneos o desactualizados.

En el plano operativo, los pilotos demostraron la necesidad inherente de un programa de capacitación básica para los entrevistadores, enfocado en el uso efectivo de la interfaz y los protocolos de recolección de datos; la ausencia de dicha formación podría constituir una barrera significativa para la adopción inicial y la fiabilidad de los datos de entrada. Finalmente, se reconoce una limitación estratégica en la integración sistémica: la interoperabilidad automática y bidireccional con sistemas de gestión académica institucionales (como SIGENU) permanece como un desarrollo pendiente en la arquitectura actual. Esta brecha limita la actualización automatizada de los listados maestros de estudiantes y exestudiantes, manteniendo un componente manual en el flujo de datos que afecta a la exhaustividad y actualidad del análisis.

Desde la perspectiva teórico-metodológica, la investigación aporta evidencia que valida la aplicación del paradigma Extracción, Transformación y Carga (ETL) en contextos de migración de datos históricos educativos complejos y poco estructurados. Los procedimientos implementados demostraron la viabilidad técnica de preservar la semántica original y las

relaciones contextuales de los registros históricos, sin comprometer la integridad referencial y la normalización del modelo de datos relacional posterior (Singu, 2022; Khan y *et al.*, 2024). En el ámbito práctico y de transferencia, la solución desarrollada establece un patrón arquitectónico y un modelo de datos reutilizable, los cuales pueden ser adaptados para futuros módulos de integración o para el análisis de otras problemáticas educativas. Este marco no solo facilita la realización de análisis longitudinales más robustos sobre la deserción, sino que también proporciona la infraestructura de datos necesaria para el diseño y evaluación de estrategias de retención estudiantil más precisas, basadas en evidencia y contextualizadas. Los hallazgos obtenidos demuestran una coherencia teórica significativa con el corpus investigativo sobre deserción en educación superior, consolidando y especificando evidencia reportada en contextos afines. En primer lugar, se corrobora la tesis de la vulnerabilidad académica previa, al validar estadísticamente que las deficiencias en competencias base— particularmente en el razonamiento lógico-matemático— constituyen un predictor sólido de riesgo de deserción en carreras de ingeniería, confirmando así hallazgos clave de estudios como el de Lázaro y *et al.* (2020). En segundo lugar, los resultados refinan la comprensión del impacto de la transición universitaria, al cuantificar y desagregar cómo factores institucionales y experiencias específicas del primer año (adaptación pedagógica, integración social, claridad vocacional) operan como desencadenantes críticos en contextos técnico-profesionales, en línea con lo documentado por Cortes *et al.* (2019). Esta investigación opera como un caso de validación empírica para los marcos conceptuales latinoamericanos que enfatizan la naturaleza multidimensional y sistémica del abandono universitario (Vega y Alfonso, 2018). Los patrones complejos identificados por la plataforma no solo ilustran esta multifactorialidad, sino que argumentan a favor de su relevancia operativa: demuestran que las intervenciones unidimensionales son insuficientes. Por tanto, los resultados sustentan y dan cuerpo práctico a la necesidad postulada en la literatura de implementar soluciones integrales y tecnológicamente habilitadas —como la aquí desarrollada— capaces de capturar, analizar y responder a esta complejidad causal de manera simultánea y contextualizada.

Conclusiones

La solución tecnológica aportada aborda pedagógicamente la gestión de la deserción estudiantil como un fenómeno central para la calidad educativa. Al informatizar integralmente el ciclo de recolección, análisis y presentación de datos, la plataforma trasciende una mera

Hacia una gestión educativa basada en datos: plataforma para el análisis de la deserción estudiantil en la CUJAE

optimización operativa para constituirse en una herramienta de construcción de conocimiento institucional. Esto permite que docentes, tutores y gestores transiten desde una comprensión intuitiva o anecdótica del problema hacia una toma de decisiones pedagógicas basada en evidencia, fundamentando así el diseño de estrategias de retención e intervención más oportunas, contextualizadas y efectivas.

La validación de la solución evidenció avances transformadores en tres dimensiones clave: eficiencia, fiabilidad y capacidad analítica. La migración exitosa de 419 registros históricos mediante un proceso ETL estructurado no solo demostró la robustez técnica, sino también la capacidad de preservar y resignificar datos pedagógicos previos. La eliminación total de errores de transcripción, la reducción del tiempo de procesamiento en más de un 98% y la funcionalidad offline con sincronización, resuelven limitaciones críticas del método manual y garantizan la confiabilidad y accesibilidad de la información, condiciones esenciales para cualquier análisis educativo serio.

La implementación de métodos no paramétricos en el análisis de datos sobre deserción ha demostrado poseer un valor pedagógico distintivo, al permitir transformar información subjetiva y cualitativa —como las percepciones recogidas en encuestas— en evidencia cuantificable y accionable para la mejora de la práctica educativa. Estos métodos han funcionado como herramientas de traducción pedagógica: el test de Kruskal-Wallis no solo ordenó carreras según la percepción docente, sino que generó un ranking diagnóstico que puede priorizar intervenciones institucionales. El test de Friedman trascendió la mera descripción estadística para fundamentar decisiones gerenciales concretas, señalando la urgente necesidad de reformar los procesos de integración estudiantil y la disponibilidad de materiales didácticos como áreas críticas de mejora. El test de Wilcoxon aportó una evidencia tranquilizadora pero pedagógicamente relevante al demostrar la ausencia de diferencias significativas entre la percepción de apoyo y la de adaptación en el acompañamiento estudiantil, lo cual valida la coherencia percibida en dicha estrategia formativa.

En conjunto, estos hallazgos refuerzan la premisa de que el análisis de datos en educación debe servir, ante todo, para iluminar procesos y fundamentar decisiones pedagógicas. Los métodos no paramétricos se revelan, así como aliados metodológicos particularmente valiosos para interpretar la complejidad humana de los fenómenos educativos, transformando percepciones y experiencias subjetivas en insumos robustos para el diseño, la evaluación y la mejora continua de las prácticas institucionales orientadas a la retención y el éxito estudiantil.

La investigación logró una integración sistémica entre el diagnóstico de necesidades pedagógicas, la selección de tecnologías apropiadas y un diseño técnico robusto. Como núcleo de este aporte, se desarrolló e integró un módulo de análisis pedagógico que, mediante técnicas de minería de datos (algoritmo Apriori en KNIME), identificó ocho patrones de asociación entre causas de baja y tipos de abandono. Este componente transforma los datos en conocimiento estratégico, permitiendo descubrir relaciones complejas y otorgando una profundidad diagnóstica inédita al fenómeno.

La solución desarrollada cierra efectivamente la brecha identificada en el estado del arte, al integrar tecnologías modernas con un enfoque de validación flexible y pedagógicamente orientado. Aunque persisten limitaciones propias de una primera iteración —como la necesidad de un programa de alfabetización digital para usuarios y la interoperabilidad pendiente con sistemas institucionales (SIGENU)—, el sistema ha demostrado ser escalable, reproducible y de alto valor educativo, constituyendo un modelo replicable para fortalecer la gestión educativa basada en datos.

Este proyecto no solo resuelve un problema operativo en la CUJAE, sino que modela un camino viable para la transformación digital con sentido pedagógico. Sus implicaciones trascienden el caso específico, demostrando cómo la tecnología puede ser un catalizador para una gestión educativa más inteligente y humana. Como trabajo futuro, se proyecta: (1) implementar la interoperabilidad total con sistemas de gestión estudiantil, (2) enriquecer el módulo analítico con técnicas de análisis predictivo y segmentación avanzada, y (3) escalar y adaptar la plataforma a otras universidades, consolidando así un aporte sustantivo a la modernización del sistema educativo superior cubano

Referencias bibliográficas

- Bautista-Díaz, M. L., Victoria-Rodríguez, E., Vargas-Estrella, L. B. y Hernández-Chamosa, C. (2020). Pruebas estadísticas paramétricas y no paramétricas: su clasificación, objetivos y características. *Educación y salud boletín científico Instituto de Ciencias de la Salud Universidad Autónoma del estado de Hidalgo*, 9(17), 78-81.
<https://doi.org/10.29057/icsa.v9i17.6293>
- Borges, G. A., Dos Santos Pedro, C. F., Santos Dos Anjos, J. C., Rodrigues, A., Boavida, F. y Sá Silva, J. (2025). A platform for early class dropout prediction of university students. *IEEE Access*, 13, 109116-109133.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11045350>

- Cortés-Cáceres, S., Álvarez, P. y Castillo, L. (2019). Deserción universitaria: la epidemia que aqueja a los sistemas de educación superior. *Revista Perspectiva*, 20(1), 13-25. <https://www.academia.edu/download/123274449/564.pdf>
- Da0, P. B. (2022). On Wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. *Applied energy*, 318, 119209. https://www.researchgate.net/profile/PhongDao/publication/360399714_On_Wilcoxon_rank_sum_test_for_condition_monitoring_and_fault_detection_of_wind_turbines/links/6274c206b1ad9f66c8a56b63/On-Wilcoxon-rank-sum-test-for-condition-monitoring-and-fault-detection-of-wind-turbines.pdf
- Del Bonifro, F. Gabrielli, M. Lisanti, G. y Zingarro, S. P. (2020). Student dropout prediction. In Proceedings of the Conference on Artificial Intelligent in Education, Lecture Notes in Computer Science, 12163, 129-140. https://link.springer.com/content/pdf/10.1007/978-3-030-52237-7_11.pdf
- Dutt, M. A., Ismail, M. A., y Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005. <https://doi.org/10.1109/ACCESS.2017.2654247>
- Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R. y Boykova, K. (2023). Data-driven decision making in higher education institutions: State-of-play. *International Journal of Advanced Computer Science and Applications*, 14(6), 397-405. <https://pdfs.semanticscholar.org/2568/e0ded66a292d2fb0b92fd00092fdb05eb9f7.pdf>
- González Santos, A. I., Vega Cruz, G. M. y Hernández González, A. (2025). Contribución a la mitigación del abandono universitario desde la práctica pedagógica de un curso introductorio. *Revista Referencia Pedagógica*, 13(1), 129-144. <http://scielo.sld.cu/pdf/rp/v13n1/2308-3042-rp-13-01-129.pdf>
- Hernández González, A. y Ansola Hazday, E. (2024). *¿Es vulnerable un estudiante que accede a carreras de ingeniería con deficiencias en las matemáticas?* [Ponencia]. RELME 17, Costa Rica.
- Hernández González, A., Martínez Girón, R. C., Ansola Hazday, E. y García Ayala, J. (2023). Estudiantes en condiciones de vulnerabilidad en la CUJAE: una aproximación a su conceptualización. *Revista Referencia Pedagógica*, 11(1), 120-137. http://scielo.sld.cu/scielo.php?pid=S2308-30422023000100120&script=sci_arttext&lng=en

- Hollander, M., Wolfe, D. A.; Chicken, E. (2014). Nonparametric statistical methods. Third Edition. John Wiley & Sons.
- Kayumova, M., Akbarova, S., Azizova, M., Hojiyeva, I., Karimova, F. y Makhmudov, F. (2024). Data Driven Teaching and Real Time Decision Making in Education Management. In *Proceedings of the 8th International Conference on Future Networks & Distributed Systems*, 756-763 <https://doi.org/10.1145/3726122.372623>
- Khan, B., Jan, S., Khan, W. y Chughtai, M. (2024). An overview of ETL Techniques Tools, Processes and Evaluations in Data Warehousing. *Journal on Big Data*, 6, 1-15. <https://media.proquest.com/media/hms/PFT/1/Rv2Wc?s=aeHmO967IMxUYxAzlr%2BEgsP1T4%3D>
- Lázaro Álvarez, N., Callejas, Z. y Griol, D. (2020). Factores que inciden en la deserción universitaria en carreras de perfil Ingeniería Informática. *Revista Fuentes*, 22(1), 105-106. <https://revistascientificas.us.es/index.php/fuentes/article/view/9262/10422>
- Liu, J. y Xu, Y. (2022). T-Friedman Test: A New Statistical Test for Multiple Comparison with an Adjustable Conservativeness Measure. *Int J Comput Intell Syst* 15, 29. <https://doi.org/10.1007/s44196-022-00083-8>
- Mckight, P. E. y Najab, J. Kruskal-wallis test (2010). The corsini encyclopedia of psychology. <https://doi.org/10.1002/9780470479216.corpsy0491>
- Niyogisubizo, J. Liao, L. Nziymva, E., Murwabashyaka, E. y Nshimyumikiza, P.C. (2022). Predicting Student's dropout university classes using two-layer ensemble machine learning approach: a novel stacked generalization. *Computer on Education Artificial Intelligence*, 3, 100066. <https://doi.org/10.1016/j.caeai.2022.100066>
- Nwokeyji, J. C. y Matouu, R. (2021). A systematic literature review on Big Data Extraction, Transformation and Loading (ETL). *Intelligent Computing*, 284. 308-324. Springer, Cham. https://doi.org/10.1007/978-3-030-80126-7_24
- Quimiz-Moreira, M., Delgadillo, R. Párraga-Alava, J., Macalau, N. y Mauricio, D. (2025). Factors, prediction, explainability and simulating university dropout through machine learning: a systematic review, 2012/2024. *Computation*, 13, 198. <https://doi.org/10.3390/computation13080198>
- Rodríguez, C. y Vega, G. (2020). *Diseño de un proyecto de investigación: estrategia para incrementar la retención en la CUJAE* [Ponencia]. X Congreso Latinoamericano de Ingeniería (CLABES).

- Singu, S. K. (2022). ETL Process Automatization: Tools and Techniques. *ESP Journal of Engineering & Technology Advancement*, 2(1), 74-85. https://www.researchgate.net/profile/Santosh-Kumar-Singu/publication/386874870_ETL_Process_Automation_Tools_and_Techniques/links/675a3011951ca355613eac3b/ETL-Process-Automation-Tools-and-Techniques.pdf
- Valencia Camacho, C. y Almeida Delgado, V. (2024). La tecnología en la gestión educativa. *Revista de Investigación Latinoamericana en Competitividad Organizacional*, 6(23), 73-79. https://ojs.eumed.net/rev/index.php/rilco/article/view/14058va/14058_pdf
- Vega Cruz, G. y Alfonso, I. (2018). Observatorio para la retención universitaria en la CUJAE. *Memorias del VIII CLABES*, 927-935. <https://revistas.utp.ac.pa/index.php/clabes/article/view/1987/3001>
- Vega Cruz, G., Oliva Pereira, M. D., Ansola Hazday, E.R., Enríquez Hernández, D. y Hernández González, A. (2025). Una acción estratégica que contribuye a la identificación del riesgo de abandono de la universidad. *Referencia Pedagógica*, 13 (3), 486. <https://rrp.cujae.edu.cu/index.php/rrp/article/view/477/515>

Contribución de autoría

Javier Molina Sánchez: desarrollador del software, *interpretación y análisis de datos* y revisión del artículo

Anaisa Hernández González: líder del proyecto de investigación, concepción del software, *interpretación y análisis de datos* y redacción del artículo.

Gilda María Vega Cruz: autora de la idea original, concepción del trabajo científico, *interpretación y análisis de datos* y revisión del artículo.

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses. Todos los autores del artículo declaramos que estamos de total acuerdo con lo escrito en este informe y aprobamos la versión final.

Autores

Javier Molina Sánchez. Estudiante de Ingeniería Informática, Universidad Tecnológica de La Habana José Antonio Echeverría, Cuba,

Anaisa Hernández González. Doctora, Profesora Titular, Vicerrectora, Universidad Tecnológica de La Habana José Antonio Echeverría, Cuba.

Gilda María Vega Cruz. Doctora, Profesora Titular, Metodóloga de la Dirección de Formación del Profesional, Universidad Tecnológica de La Habana José Antonio Echeverría, Cuba,

