

Predictores claves del abandono escolar en Paraguay: Un análisis de datos masivos

Key Predictors of School Dropout in Paraguay: A Big Data Analysis

<https://doi.org/10.4438/1988-592X-RE-2025-409-686>

Jorge Daniel Mello-Román

<https://orcid.org/0000-0003-2821-7538>

Universidad Nacional de Concepción

Ricardo Daniel Escobar-Torres

<https://orcid.org/0000-0001-8764-6784>

Universidad Nacional de Concepción

Marina Segura

<https://orcid.org/0000-0002-6068-960x>

Universidad Complutense de Madrid

M. Covadonga de la Iglesia Villasol

<https://orcid.org/0000-0001-5092-9932>

Universidad Complutense de Madrid

Salvadora Giménez

<https://orcid.org/0000-0003-0915-9328>

Universidad Nacional de Concepción

Adolfo Hernández

<https://orcid.org/0000-0003-1078-2328>

Universidad Complutense de Madrid

Julio César Mello-Román

<https://orcid.org/0000-0002-3698-4043>

Universidad Nacional de Concepción

Pastor E. Pérez Estigarribia

<https://orcid.org/0000-0003-3143-0751>

Universidad Nacional de Asunción

Resumen

El abandono escolar es un problema estructural en América Latina con profundas implicaciones en el desarrollo social y económico. Este estudio analiza los factores asociados al abandono escolar en los años previos a la finalización de la educación secundaria (denominada Educación Media en Paraguay), utilizando datos administrativos del Registro Único del Estudiante (RUE), el sistema de gestión de datos educativos del Ministerio de Educación y Ciencias, durante el período 2017-2023. A través de un enfoque cuantitativo, se realizaron análisis descriptivos y se aplicaron modelos de aprendizaje automático con el objetivo de identificar patrones de abandono escolar y predecir su riesgo. Se examinaron 706.785 registros estudiantiles, considerando variables sociodemográficas, académicas e institucionales. Los resultados evidenciaron un incremento en las tasas de deserción entre 2019 y 2020, coincidiendo con el impacto de la pandemia de COVID-19. Se observaron diferencias significativas según género, especialidad educativa y ubicación geográfica, identificándose la sobreedad y la repetición como factores críticos del abandono escolar. Los estudiantes matriculados en el turno nocturno y en programas de formación profesional presentaron las tasas de deserción más elevadas. En términos predictivos, la regresión LASSO mostró el mejor desempeño, logrando un equilibrio óptimo entre precisión y sensibilidad en la identificación de estudiantes en riesgo. Estos hallazgos subrayan la importancia del análisis de datos masivos y la aplicación de modelos avanzados para fortalecer las políticas de retención escolar y diseñar estrategias de intervención temprana basadas en evidencia. No obstante, persisten desafíos relacionados con la calidad y cobertura de los datos educativos, la necesidad de explorar metodologías emergentes de inteligencia artificial e integrar factores psicosociales y económicos para una comprensión integral de la deserción escolar y sus determinantes.

Palabras clave: Abandono escolar, Educación secundaria, Nivel socioeconómico, Ciencia y análisis de datos, Predictores.

Abstract

School dropout constitutes a structural challenge in Latin America, carrying profound implications for social and economic development. This study examines the factors associated with school dropout in the years preceding the completion of secondary education (known in Paraguay as Educación Media), utilising administrative data from the Registro Único del Estudiante (RUE), the educational data management system of the Ministry of Education and Science, for the period 2017-2023. A quantitative approach was employed,

encompassing descriptive analyses and the application of machine learning models to identify dropout patterns and predict dropout risk. The study analyzed 706,785 student records, incorporating sociodemographic, academic, and institutional variables. The findings indicate a significant increase in dropout rates between 2019 and 2020, coinciding with the onset of the COVID-19 pandemic, with notable differences observed across gender, educational pathway, and geographic location, where grade overage and grade repetition emerged as critical determinants of dropout. Students enrolled in night shift and vocational training programs exhibited the highest dropout rates. In predictive terms, LASSO regression demonstrated the best performance, achieving an optimal balance between precision and sensitivity in identifying at-risk students. These results highlight the importance of leveraging extensive data analysis and advanced modelling techniques to strengthen school retention policies and develop evidence-based early intervention strategies. However, challenges remain concerning the quality and comprehensiveness of educational data, the need to explore emerging artificial intelligence methodologies, and the integration of psychosocial and economic factors to achieve a holistic understanding of school dropout and its determinants.

Key words: Dropouts, Secondary Education, Socioeconomic Background, Data Science, Predictor Variables.

Introducción

El abandono escolar (también denominado abandono educativo temprano o deserción escolar) es un fenómeno complejo que impacta significativamente el desarrollo educativo y social (Zengin, 2021). A nivel global, millones de estudiantes abandonan la escuela cada año, lo que plantea desafíos para el desarrollo sostenible. En América Latina, el abandono escolar supera el 30% en algunos países, con un 41,4% de la población activa mayor de 15 años sin completar la secundaria (Comisión Económica para América Latina y el Caribe [CEPAL], 2024), agravándose con la pandemia. En contraste, la Unión Europea ha reducido el abandono al 9,5% en 2024, aunque España aún registra un 13% (EUROSTAT, 2025).

En Paraguay, el abandono escolar ha recibido creciente atención, especialmente en el marco de las políticas del Ministerio de Educación y Ciencias (MEC). Según la Encuesta Permanente de Hogares Continua 2023 del Instituto Nacional de Estadística (INE, 2024), el 98,1% de la población de 10 a 14 años asistió a un centro educativo, cifra que cae al 71,4% en el grupo de 15 a 19 años, con niveles menores en áreas rurales (65,5%) frente a urbanas (75,2%).

Este estudio se centra en el abandono escolar durante los años previos

a la finalización de la educación secundaria, una etapa fundamental para la formación del capital humano y la preparación para el mercado laboral y la vida cívica. El abandono en este nivel tiene consecuencias duraderas: a nivel individual, se asocia con mayores tasas de desempleo y empleos de menor calidad, mientras que a nivel social perpetúa ciclos de pobreza y desigualdad.

La creciente disponibilidad de datos masivos y los avances en modelos de aprendizaje automático han demostrado ser herramientas efectivas para predecir el abandono y el rendimiento académico (Krüger et al., 2023). No obstante, en América Latina son escasos los estudios que evalúan estas tecnologías en niveles preuniversitarios (Rodríguez et al., 2023; Smith y Gutiérrez, 2022), y la infraestructura y calidad de datos siguen siendo desafíos (Hernández-Leal et al., 2021).

Actualmente, los sistemas de gestión de datos educativos (SIGED) juegan un papel crucial en el seguimiento académico y la toma de decisiones basadas en datos, aunque enfrentan retos de interoperabilidad, calidad y cobertura (Arias et al., 2021). La explotación de estos datos masivos ofrece una oportunidad para mejorar la eficiencia de las políticas educativas.

En Paraguay, el Registro Único del Estudiante (RUE) es la principal herramienta de gestión de datos del sistema educativo, desarrollado por el MEC. Este sistema relacional, operativo desde 2017, identifica de forma única a cada estudiante y rastrea su trayectoria académica, integrando variables como historial académico, datos familiares y condiciones socioeconómicas (INE, 2024), aunque presenta limitaciones en variables contextuales.

El presente estudio tiene como objetivo analizar los patrones de abandono escolar en la educación secundaria (denominada Educación Media en Paraguay) utilizando datos del RUE durante el periodo 2017-2023. Se plantean las siguientes preguntas: ¿Cuáles son los principales patrones de abandono en la Educación Media en Paraguay? ¿Qué variables sociodemográficas e institucionales inciden críticamente en el abandono? ¿Qué técnicas de análisis y predicción ofrecen mayor capacidad para identificar tempranamente a los estudiantes en riesgo?

La respuesta a estas preguntas generará un diagnóstico basado en evidencia para fortalecer las políticas de retención escolar en Paraguay y contribuir al debate global sobre la reducción de la desigualdad educativa, ofreciendo insumos comparables para otros países y promoviendo el uso de datos masivos en la mejora de los sistemas educativos y en la formulación de estrategias inclusivas que garanticen el acceso y la permanencia de los estudiantes.

Marco Teórico

El abandono escolar en la educación secundaria es un fenómeno complejo que afecta a América Latina y Europa, con repercusiones sociales, económicas y culturales. Pese a ciertos avances, continúa siendo un desafío para ambas regiones (CEPAL, 2024; European Education Area, 2022). Se define como el abandono prematuro de los estudios antes de finalizar la secundaria —etapa que, en España, se corresponde principalmente con la formación postobligatoria—, lo cual limita oportunidades futuras y refuerza ciclos de pobreza y exclusión. Sus factores explicativos son múltiples y se desarrollan gradualmente, por lo que su prevención requiere intervenciones sostenidas y personalizadas en función del grado de vulnerabilidad del estudiante (Khurram et al., 2023; Hernández y Alcaraz, 2018). Ramsdal y Wynn (2022) recomiendan la psicología positiva y el trabajo en grupo para evitarla, mientras que Nait (2021) resalta el gran esfuerzo necesario en zonas rurales.

Factores explicativos del abandono escolar

La pobreza es uno de los factores más relevantes en América Latina: muchos estudiantes deben abandonar la escuela para trabajar y sostener a sus familias. En Paraguay, más del 66% de quienes abandonan lo hacen por razones económicas. El trabajo infantil también es frecuente, afectando a la asistencia y el rendimiento en niños y adolescentes. En Europa existen retos similares, pero los sistemas de protección social han ayudado a atenuarlos. Aun así, en países como España la tasa de abandono escolar alcanza un 13% en 2024 (Ministerio de Educación, Formación Profesional y Deportes [MEFPD], 2025).

El bajo rendimiento académico y la falta de metodologías de enseñanza efectivas constituyen otro factor determinante. En América Latina, la insuficiente actualización de los programas y la desconexión con las realidades de los estudiantes provocan desmotivación, al no responder a sus intereses, vocaciones ni contextos culturales. En varios países de la región, incluido Paraguay, más de la mitad de la población de 15 años no alcanza el nivel mínimo de competencias en resolución de problemas matemáticos (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2022, p. 92).

El entorno familiar y social también incide en el abandono escolar. En América Latina, factores como el embarazo adolescente, las uniones tempranas, la pobreza y el trabajo infantil afectan significativamente la permanencia

escolar, especialmente entre las adolescentes. Además, la desescolarización es mayor entre jóvenes indígenas y afrodescendientes (CEPAL, 2024). En Europa, sin embargo, la educación sexual redujo el abandono por embarazo, salvo en algunas comunidades migrantes (European Education Area, 2022)

Las características institucionales también pesan. En América Latina, persisten importantes deficiencias de infraestructura y brechas digitales. En Paraguay, menos del 50% de las escuelas secundarias cuenta con conectividad para fines pedagógicos (UNESCO, 2022, p. 80). Esto, sumado a la brecha entre la educación pública y la privada, que favorece a estas últimas, con tasas de abandono menores. Aunque Europa presenta condiciones más favorables, persisten disparidades entre países y regiones (EUROSTAT, 2025).

Consecuencias del abandono escolar

A nivel individual, no concluir la educación secundaria limita las posibilidades de acceder a empleos de calidad y perpetúa condiciones de pobreza y exclusión. En América Latina, el abandono escolar está vinculado con el trabajo infantil y la inserción temprana en el mercado laboral informal, lo que afecta el derecho a una educación de calidad y reduce las oportunidades de ejercer una ciudadanía plena (CEPAL, 2024).

En el plano social, el abandono escolar profundiza las desigualdades estructurales y refuerza el ciclo de reproducción intergeneracional de la pobreza. En América Latina, mientras que el 85% de los jóvenes del quintil con mayores ingresos logra completar la secundaria, solo el 44% lo consigue en el quintil más pobre (UNESCO, 2022, p. 17). Esta brecha impacta directamente en el acceso a oportunidades y supone mayores exigencias para los sistemas de asistencia y protección social, incrementando la demanda de apoyo público (CEPAL, 2024).

Desde una perspectiva económica, el abandono escolar socava la productividad y la competitividad de los países. Según la CEPAL (2024) cada año extra de escolaridad eleva los ingresos entre un 7% y un 10%. A nivel macro, los países con altas tasas de abandono tienen menor crecimiento del PIB, reducirla en un 10% podría aumentar el PIB en 0,5% anual. La educación secundaria proporciona habilidades esenciales para acceder a empleos calificados; quienes abandonan tienen menos oportunidades en el mercado formal, ingresando a empleos precarios y mal remunerados. En Paraguay, alrededor del 60% de los jóvenes que abandonan la Educación Media provienen de hogares vinculados a actividades económicas informales (Ministerio

de Educación y Ciencias [MEC], 2013, p. 61). Mientras en América Latina el crecimiento económico se ve limitado por la alta informalidad laboral, en Europa, donde el abandono escolar es menor, las economías han desarrollado sectores tecnológicos y basados en el conocimiento (European Education Area, 2022).

Datos masivos en estudios sobre abandono escolar

El uso de datos masivos (Big Data) en educación impulsa la detección y prevención del abandono escolar al procesar gran volumen de datos en tiempo real (Amaya-Amaya et al., 2020). Suele preceder al abandono, el absentismo o la repetición del curso (Montero-Sieburth y Turcatti, 2022; Ajjawi et al., 2020). En este sentido, el enfoque de analítica de datos, las técnicas de aprendizaje automático (machine learning) y la minería de datos educativos (Flores, 2025; Serrano et al., 2024) han demostrado eficacia para detectar el abandono, y su precisión depende de la calidad y diversidad de las variables (Kocsis y Molnár, 2024).

Las variables que explican el abandono abarcan ámbitos personales, socioeconómicos, académicos e institucionales (Alyahyan y Düşteğör, 2020; Alladatin et al., 2023), y ningún factor por sí solo resulta decisivo. Gutiérrez-de-Rozas et al. (2023) resaltan la orientación académico-profesional y aspectos personales como desencadenantes, mientras que Zapata-Medina et al. (2024) y otros hallan que las variables académicas suelen ser las más predictivas (Segura et al., 2022; Ortiz-Lozano et al., 2023). En Paraguay, Insfrán-Coronel et al. (2024) identifican el papel crítico de la pobreza y el trabajo infantil. Asimismo, Abideen et al. (2023), junto con Venkatesan y Mappillairaju (2023), corroboran la importancia de la minería de datos para ubicar áreas con alta incidencia, guiando intervenciones más eficaces.

En la actualidad la mayoría de los estudios sobre datos masivos y modelos predictivos del abandono se enfocan en la educación superior, con una brecha importante con relación a su implementación en niveles previos (Rodríguez et al., 2023, Smith y Gutiérrez, 2022). Métodos de ensamble de árboles de decisión, como Random Forest, destacan por su eficacia predictiva, mientras que técnicas como Regresión Logística siguen siendo recurridas por su interpretabilidad. Modelos más complejos, como Redes Neuronales, son menos frecuentes (Venkatesan y Mappillairaju, 2023). En América Latina la falta de datos estructurados y accesibles aún limitan su implementación, junto con desafíos tecnológicos, de calidad y cobertura (Hernández-Leal et al., 2021; Arias et al., 2021).

Método

Fuente de datos

Los datos provienen del Registro Único del Estudiante (RUE), administrado por el Ministerio de Educación y Ciencias (MEC) de Paraguay. El RUE centraliza información sobre la trayectoria educativa de los estudiantes a nivel nacional, facilitando el análisis de factores relacionados con la permanencia y graduación escolar. Las instituciones reportan periódicamente al MEC, y esta información se combina con otras bases oficiales para mejorar su precisión (INE, 2024).

Este estudio se centra en 706.785 estudiantes de Educación Media de Paraguay, distribuidos en 9.051 instituciones, abarcando diversas especialidades. Las variables incluyen características sociodemográficas, institucionales y académicas. Para el año 2017, se dispone además de variables contextuales relacionadas con la salud, el entorno familiar, la vivienda y las condiciones socioeconómicas, aunque con algunas limitaciones en su reporte. En los años posteriores, la captura de estas variables no se continuó de forma sistemática, lo que convierte a 2017 en la única cohorte con información contextual, motivando la concentración del análisis en dicho periodo.

La base de datos, en formato relacional, fue gestionada con PostgreSQL, lo que garantiza la integridad y eficiencia en el manejo de grandes volúmenes de información. En ningún caso se accedió a datos que permitan la identificación nominal de los estudiantes (por ejemplo, nombres o DNI), garantizando así la confidencialidad.

Procesamiento de datos

Se realizó una revisión general de la consistencia de la información, identificando respuestas contradictorias, duplicadas o con errores sistemáticos. A partir de ello, se aplicaron procesos de limpieza y estandarización para unificar los datos conflictivos y eliminar aquellos valores poco representativos. Asimismo, se evaluó la pertinencia de cada variable en relación con el fenómeno del abandono escolar, excluyendo aquellas que no presentaban una relación clara o mostraban inconsistencias en su reporte, de modo a maximizar la calidad del análisis y evitar la incorporación de información que pudiera sesgar los resultados.

La construcción de variables derivadas se fundamentó en el segui-

miento longitudinal de los estudiantes. Se definió como abandono escolar a aquellos alumnos que, tras su última inscripción (entre 2017 y 2021), ni se graduaron ni volvieron a matricularse en como mínimo dos años consecutivos. Por otro lado, la graduación escolar se estableció como la situación en la que el estudiante alcanza el último año de su ciclo formativo sin matriculaciones posteriores. Estas definiciones fueron esenciales para el análisis posterior.

Enfoque del estudio

Este trabajo adopta un enfoque cuantitativo, combinando análisis descriptivo del abandono escolar, explorando frecuencias y tasas segmentadas por variables sociodemográficas, institucionales y temporales, y modelos predictivos que utilizan técnicas de aprendizaje automático para anticipar la probabilidad de abandono, considerando información del estudiante, la institución y el contexto socioeconómico.

El análisis descriptivo se planteó con el objetivo de caracterizar la población de estudio y examinar la distribución del abandono escolar en función de variables sociodemográficas, institucionales y geográficas. Para ello, se estimaron frecuencias y tasas de abandono segmentadas por género, zona, tipo de institución y modalidad educativa, que se presentan en tablas, gráficos y mapas. Esta aproximación inicial permite identificar los principales patrones del abandono escolar en la educación media paraguaya, sirviendo como base para las siguientes etapas de análisis.

El enfoque predictivo se desarrolló mediante el análisis de una cohorte de estudiantes que ingresaron al primer curso de la Educación Media en Paraguay en 2017, con el fin de evaluar los factores que inciden en aquellos que no culminaron el nivel educativo. Tras procesar y depurar el conjunto de datos, se llevó a cabo una rigurosa selección de atributos y se implementaron diversos modelos predictivos. Los procesos, modelos y técnicas empleados se detallan en la siguiente sección.

Estudio de Caso: Cohorte 2017

Se analizó una cohorte de estudiantes que ingresaron al primer curso de la Educación Media en Paraguay en el año 2017, para evaluar los factores asociados al abandono escolar y examinar la trayectoria académica de los alumnos. La variable objetivo, denominada “Abandono escolar”, se define a

partir de la interrupción de la matrícula en etapas posteriores a la inscripción inicial, permitiendo identificar de forma clara a aquellos estudiantes que no continúan en el sistema. El conjunto de datos integra variables de diversas categorías, tales como características demográficas (sexo, edad y sobriedad), historial académico (cursos repetidos, tipo de especialidad educativa y turno), rasgos institucionales (tipo de gestión y ubicación geográfica), condiciones socioeconómicas (nivel educativo y ocupación de padres o tutores, características del hogar) y aspectos relacionados con el acceso y la movilidad (tiempo y medio de transporte utilizado para llegar a la escuela).

El primer paso del proceso analítico consistió en llevar a cabo un análisis exploratorio para identificar variables con altos porcentajes de datos faltantes. Se eliminaron aquellas variables que presentaban porcentajes elevados de ausencias, dada su poca fiabilidad para contribuir al análisis. En cuanto a los registros individuales, la base inicial contenía 87.361 observaciones, las cuales fueron depuradas mediante la aplicación de umbrales sucesivos de eliminación por valores faltantes. Comenzando con un umbral del 80% y reduciendo progresivamente hasta llegar a registros completos (0% de datos faltantes), se obtuvo una base final de 63.290 observaciones. Durante este proceso, se mantuvo el equilibrio en la variable “Abandono escolar” (entre el 15% y 12% de abandono), lo que respalda la representatividad de la muestra y garantiza su idoneidad para la implementación de modelos analíticos.

Los identificadores institucionales fueron excluidos del análisis, previa comprobación de que esta decisión no afectaba el desempeño de los modelos predictivos. Aunque se determinó un Coeficiente de Correlación Intraclass (ICC) de aproximadamente 33%, lo que sugiere que una parte considerable de la variabilidad en el abandono escolar se debe a diferencias entre instituciones, su exclusión evita que los modelos se ajusten a patrones específicos y garantiza su capacidad de generalización a nuevos contextos. Además, ante el marcado desbalance de la variable “Abandono escolar”, se implementó, como paso previo a la generación de los modelos predictivos, una técnica de submuestreo (undersampling) sobre la clase mayoritaria, de modo que el número de registros se equilibrara adecuadamente con la clase minoritaria, asegurando así la representatividad de los datos críticos.

Selección de Atributos

La adecuada selección de atributos es crucial para reducir el sobreajuste, garantizando que el modelo capture patrones generales y relevantes en lugar

de ajustarse a las particularidades del conjunto de datos. Para identificar de forma robusta las variables más relevantes para la predicción del abandono escolar se emplearon diversas técnicas de selección de atributos. Esto con el fin de asegurar que la selección se base en criterios diversificados y sólidos.

En primer lugar, se utilizó el enfoque de Random Forest, aplicando dos criterios de importancia: Mean Decrease Accuracy (MDA) y Mean Decrease Gini (MDG). Estos criterios permiten evaluar la contribución directa de cada variable a la capacidad predictiva del modelo, ya sea a través de la disminución en la exactitud al permutar sus valores o mediante la reducción de la impureza en los nodos del árbol. Paralelamente, se aplicó la Regresión Logística, que al estimar la relación entre las variables independientes y la probabilidad de abandono, ofrece coeficientes interpretables y facilita la identificación de predictores significativos. Complementariamente, se utilizó la Regresión LASSO, que incorpora un término de penalización L1 para forzar a cero los coeficientes de aquellas variables menos relevantes, mitigando así el riesgo de sobreajuste y favoreciendo la generalización del modelo.

Modelos Predictivos del Abandono Escolar

Con las variables seleccionadas se implementaron diversos modelos de clasificación para evaluar su capacidad predictiva en la identificación del abandono escolar, abarcando enfoques lineales, basados en árboles de decisión y técnicas clásicas de aprendizaje automático. Previamente, se realizó una revisión y ajuste de los hiperparámetros de cada modelo para maximizar su desempeño.

La Regresión Logística se empleó como modelo base, dada su alta interpretabilidad. Los coeficientes estimados mediante máxima verosimilitud permiten calcular los Odds Ratio —también conocidos como “odds” o razón de probabilidad—, lo que facilita evaluar la influencia de cada predictor en el contexto del abandono escolar. Asimismo, se utilizó la Regresión LASSO, que incorpora regularización mediante un término de penalización L1, para minimizar el error de ajuste. En este caso, se aplicó validación cruzada de 10-folds para determinar el parámetro óptimo λ , permitiendo descartar coeficientes de variables poco relevantes y reducir el riesgo de sobreajuste.

Adicionalmente, se implementaron algoritmos basados en árboles de decisión, agrupando Random Forest y Extreme Gradient Boosting (XGBoost). Random Forest se configuró con un número adecuado de árboles y parámetros ajustados para reducir la varianza y mejorar la estabilidad del mo-

delo, mientras que XGBoost optimizó iterativamente la precisión predictiva mediante boosting, utilizando parámetros diseñados para prevenir el sobreajuste. También se incluyeron técnicas clásicas de aprendizaje automático, como las Máquinas de Soporte Vectorial (SVM) con kernel radial, para captar relaciones no lineales, y una Red Neuronal Artificial Multicapa (ANN), en la que se estandarizaron las variables predictoras y se optimizaron los pesos mediante retropropagación. La ampliación de la variedad de modelos predictivos se fundamenta en garantizar una mayor capacidad de generalización de las conclusiones, complementando lo aportado por las técnicas de selección de variables.

Para comparar el desempeño de los modelos se emplearon métricas estándar de clasificación: exactitud (Accuracy), precisión (Precision), sensibilidad (Recall or Sensitivity), especificidad (Specificity) y F1-Score. En el contexto del abandono escolar, Accuracy mide la proporción total de predicciones correctas, ofreciendo una visión global del rendimiento del modelo. Precision indica la proporción de predicciones positivas acertadas, lo que resulta crucial para minimizar falsas alarmas en la identificación de estudiantes en riesgo. Recall o Sensitivity, es especialmente importante, ya que refleja la capacidad del modelo para detectar correctamente a los estudiantes que efectivamente abandonan, lo cual es fundamental para la implementación de intervenciones tempranas. Por otro lado, Specificity evalúa la habilidad para identificar con precisión a aquellos estudiantes que continúan en el sistema educativo, mientras que el F1-Score, al combinar Precision y Sensitivity, proporciona una medida equilibrada del rendimiento general del modelo.

Para el desarrollo del estudio se emplearon diversas herramientas tecnológicas. La base de datos RUE se gestionó mediante PostgreSQL. Para el análisis descriptivo y la modelación predictiva se utilizaron Python y R, mientras que se recurrió a Microsoft Excel para realizar cálculos exploratorios, validar datos y presentar resultados.

Resultados

El análisis descriptivo se basa en 706.785 estudiantes de Educación Media de Paraguay (370.032 femeninas y 336.753 masculinos) durante 2017-2023, evidenciando una ligera predominancia femenina. La Tabla I muestra un crecimiento en el número de matriculados desde 2017, alcanzando un pico en 2021 y una leve disminución en 2022 y 2023. Asimismo, el número de aban-

donos aumentó, con picos superiores a 21.000 casos en 2019 y 2020, coincidiendo con el inicio de la pandemia COVID-19, lo que pudo afectar la permanencia escolar. En 2021, los abandonos disminuyeron a 14.918, en línea con el retorno parcial a la presencialidad y la implementación de planes de contingencia. Además, inicialmente la tasa de abandono fue mayor en hombres (2017-2019), pero a partir de 2020 el porcentaje de abandono femenino superó al masculino, alcanzando 52,25% en 2020 y 57,47% en 2021.

TABLA I. Distribución de la Frecuencia y Porcentaje de Abandono Escolar en Educación Media por Año y Género (2017-2023).

Año	Total Inscritos (T.I.)	Frecuencia Abandono (F.A.)	% Total Abandono (% A.)	F. A. Mujeres	% A. Mujeres	F. A. Hombres	% A. Hombres
2017	87.361	13.378	15,31	5.288	39,53	8.090	60,47
2018	93.061	13.450	14,45	5.809	43,19	7.641	56,81
2019	103.981	21.065	20,26	10.064	47,78	11.001	52,22
2020	106.497	21.969	20,63	11.479	52,25	10.490	47,75
2021	107.529	14.918	13,87	8.574	57,47	6.344	42,53
2022	105.019	-	-	-	-	-	-
2023	103.337	-	-	-	-	-	-
Total	706.785	84.780		41.214	48,61	43.566	51,39

Fuente: Base de datos RUE, elaboración propia.

La Tabla II muestra que la edad promedio al inscribirse en Educación Media aumentó de 15,58 años en 2017 a 17,59 años en 2023, lo que indica un aumento en el ingreso tardío al nivel educativo. Asimismo, la edad promedio al abandonar se elevó de 16,47 años en 2017 a 23,46 años en 2021, reflejando una mayor permanencia antes de abandonar; este incremento es notable en 2020 y 2021, también coincidente con la pandemia de COVID-19. En general, las mujeres presentan una edad promedio de abandono superior a la de los hombres, con una diferencia que se acentúa progresivamente durante todo el periodo de análisis.

TABLA II. Edad Promedio al Inscribirse y al Abandonar la Educación Media por Año y Género (2017-2023)

Año	Edad Promedio al inscribirse	Edad Promedio al abandonar (E.P.A.)	E. P. A. Mujeres	E. P. A. Hombres
2017	15,58	16,47	16,18	16,65
2018	15,79	17,54	18,07	17,14
2019	16,62	20,47	22,01	19,06
2020	16,99	21,72	23,37	19,91
2021	17,4	23,46	24,72	21,78
2022	17,44	-	-	-
2023	17,59	-	-	-

Fuente: Base de datos RUE, elaboración propia.

En la Tabla III se observa que, en promedio, los estudiantes de zonas rurales tienen una tasa de abandono del 15,95% (3.692 de 23.488 inscriptos), frente al 17,39% de los estudiantes en zonas urbanas, manteniéndose estos valores de manera similar durante todo el periodo. Asimismo, se aprecian picos máximos de abandono en los años 2019 y 2020, siendo las áreas urbanas las más afectadas; solo en 2019, la cantidad de abandonos aumentó de 9.979 a 16.037, lo que representa un incremento del 60%. Para el año 2021, los patrones de abandono escolar parecen retroceder a los niveles observados al inicio del periodo de análisis.

TABLA III. Distribución de la Frecuencia y Porcentaje de Abandono Escolar en Educación Media por Año y Zona (2017-2023)

Año	T. I. Rural	T.I. Urbana	F.A. Rural	% A. Rural	F.A. Ur-bano	% A. Ur-bano
2017	23.488	63.873	3.692	15,72	9.686	15,16
2018	24.698	68.363	3.471	14,05	9.979	14,60
2019	27.558	76.423	5.028	18,25	16.037	20,98
2020	27.978	78.519	5.424	19,39	16.545	21,07
2021	28.173	79.356	3.421	12,14	11.497	14,49
2022	26.415	78.604	-	-	-	-
2023	25.031	78.306	-	-	-	-
Total	183.341	523.444	21.036	15,95	63.744	17,39

Fuente: Base de datos RUE, elaboración propia.

En la Tabla IV se observa que el abandono escolar se concentra principalmente en las instituciones oficiales, que representan más del 80% de los casos, y con una ligera disminución en 2021 (78,50%). Por otro lado,

las instituciones privadas han experimentado un incremento del abandono en términos comparativos a lo largo de todo el periodo de análisis (2017 - 2021), mientras que las instituciones privadas subvencionadas se mantienen estables con una baja tasa de abandono.

TABLA IV. Distribución de la Frecuencia y Porcentaje de Abandono Escolar en Educación Media por tipo de institución (2017-2021)

Año	F.A. Oficial	% A. Oficial	F.A. Privado	% A. Privado	F. A. Priv. Subv.	% A. Priv. Subv.
2017	11565	86,45	1.136	8,49	677	5,06
2018	11116	82,65	1.727	12,84	607	4,51
2019	17402	82,61	3.036	14,41	627	2,98
2020	18375	83,64	2.856	13,00	738	3,36
2021	11711	78,50	2.765	18,53	442	2,96
Total	70.169	82,77	11.520	13,59	3.091	3,65

Fuente: Base de datos RUE, elaboración propia.

El análisis de la Tabla V revela que el Bachillerato Científico concentra la mayor cantidad de inscripciones (343.919) y abandonos absolutos (46.554 casos, equivalente al 54,91% del total); sin embargo, su tasa de abandono relativa es moderada (13,54%), lo que indica que, a pesar de ser la especialidad más concurrida, proporcionalmente no enfrenta el mayor riesgo. En contraste, los abandonos en los bachilleratos Sin Especialidad representan el 32,97% del total (27.956 casos) y muestran una tasa de abandono relativa alta (61,77%), lo que la convierte en la especialidad más vulnerable. Asimismo, las modalidades técnicas, como el Bachillerato Técnico Industrial y el Bachillerato Técnico en Servicio, presentan bajas tasas relativas (5,48% y 7,76%, respectivamente). Por último, la Formación Profesional Vocacional, a pesar de su baja matrícula y abandono absoluto, exhibe la mayor tasa de abandono relativa (63,48%).

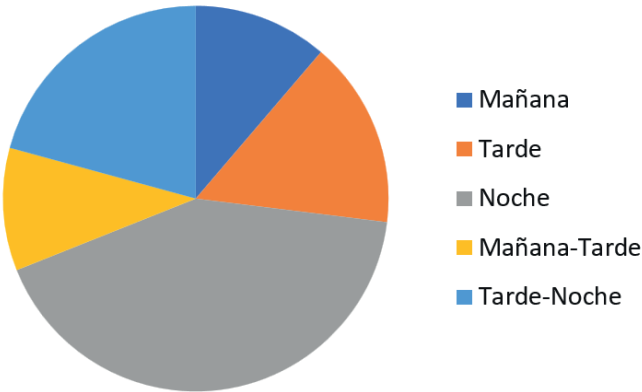
TABLA V. Distribución de la Frecuencia y Porcentaje de Abandono Escolar en Educación Media por tipo de especialidad (2017-2021)

Tipo de Especialidad	Total de Ins-criptos (T. I.)	Frecuencia Abandono (F.A.)	% Total Abandono (% A.)	% A. Rela-tivo
Bachillerato Científico	343.919	46.554	54,91	13,54
Bachillerato Técnico en Servicio	80.374	6.241	7,36	7,76
Sin Especialidad	45.257	27.956	32,97	61,77
Bachillerato Técnico Industrial	14.568	798	0,94	5,48
Bachillerato Técnico Agropecua-rio	11.797	1.635	1,93	13,86
Formación Profesional Vocacio-nal	2.514	1.596	1,88	63,48

Fuente: Base de datos RUE, elaboración propia.

El Gráfico I muestra el porcentaje de abandono relativo por turno escolar, evidenciando diferencias significativas. El turno noche registra el porcentaje más alto (42%), probablemente debido a las cargas laborales de sus estudiantes. Le siguen el turno tarde-noche (20,8%) y el turno tarde (15,7%), que también presentan un riesgo considerable. En contraste, los turnos mañana-tarde (10,3%) y mañana (11,3%) muestran los valores más bajos.

GRÁFICO I. Porcentaje de Abandono Escolar Relativo en Educación Media por turno (2017-2021)



Fuente: Base de datos RUE, elaboración propia.

En la Tabla VI se observa que los estudiantes que no han repetido ningún año en la Educación Media presentan una tasa de abandono baja (11,70%). En contraste, aquellos que han repetido al menos una vez muestran una tasa significativamente mayor (24,92%), lo que indica que la repetición de un año académico se asocia con un mayor riesgo de deserción. Asimismo, la tasa relativa se mantiene en niveles similares para los estudiantes con dos y tres cursos repetidos.

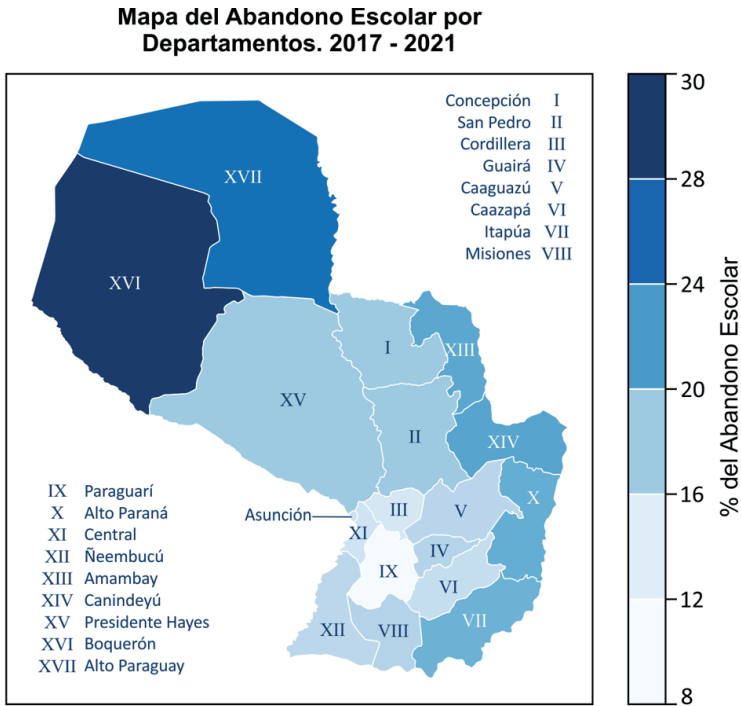
TABLA VI. Distribución de la Frecuencia y Porcentaje de Abandono Escolar en Educación Media por Nro de cursos repetidos (2017-2021)

Número de cursos repetidos	Total	No abandono	Abandono	% Abandono
0	690.499	609.743	80.756	11,70
1	14.926	11.207	3.719	24,92
2	1.214	937	277	22,82
3	133	106	27	20,30
más de 4	13	12	1	7,69

Fuente: Base de datos RUE, elaboración propia.

El Gráfico II muestra la tasa de abandono escolar por departamento en Paraguay. Se observa que los departamentos de Alto Paraguay y Boquerón registran tasas superiores al 24%, probablemente debido a barreras geográficas y condiciones socioeconómicas adversas. En contraste, departamentos como Central y Paraguari presentan tasas menores. Asimismo, se evidencia que aquellos departamentos con fronteras dinámicas con países como Brasil y Argentina registraron, en promedio, mayores tasas de abandono durante el periodo de análisis.

GRÁFICO II. Mapa del Abandono Escolar en Educación Media (2017-2021)



Fuente: Base de datos RUE, elaboración propia.

Estudio de Caso: Cohorte 2017

Se desarrolló un enfoque predictivo basado en el análisis de una cohorte de estudiantes que ingresaron al primer curso de Educación Media en Paraguay en 2017. Se realizó una rigurosa selección de atributos e implementaron diversos modelos predictivos del abandono escolar; los resultados se presentan a continuación.

Selección de Atributos

Esta sección resume las métricas obtenidas mediante Regresión Logística, Random Forest (evaluado con los criterios MDA y MDG) y Regresión LAS-SO para las cinco variables seleccionadas según los cuatro criterios. La Tabla VI presenta los valores numéricos de las métricas para las variables relevan-

tes en cada uno de los métodos, permitiendo identificar la importancia relativa de cada variable.

TABLA VII. Métricas de Selección de atributos. Variables relevantes para los métodos implementados.

Variable	Regresión Logística (Importancia)	Random Forest (MDA)	Random Forest (MDG)	Regresión LASSO (Coeficiente)
Edad (en años)	15.7	34.03	492.42	0.611
Especialidad	23.3	16.85	256.09	5.49
Turno	12.6	12.57	158.14	1.29
Departamento	6.71	8.54	390.69	2.33
Sobreedad	2.59	22.27	216.42	0.201

Fuente: Base de datos RUE, elaboración propia.

Las variables Edad en años, Especialidad, Turno, Departamento y Sobreedad se seleccionaron de forma consistente, capturando dimensiones clave del perfil estudiantil y del contexto institucional. La literatura respalda que la edad, y especialmente la sobreedad, es un determinante crítico de la permanencia escolar, ya que el rezago académico aumenta el riesgo de abandono. Además, la Especialidad, Departamento y Turno pueden estar reflejando el efecto de desigualdades socioeconómicas y culturales en la continuidad escolar.

Otras variables se eligieron moderadamente en tres métodos, como Modalidad—que clasifica el currículo en Bachillerato Científico, Técnico, de Formación Profesional y Abierta—, la Gestión de la institución (Pública, Privada o Subvencionada), y las características de la vivienda, como la disponibilidad de electrodomésticos, que reflejan el nivel socioeconómico.

Finalmente, por lo menos dos métodos coincidieron en seleccionar variables como el Sexo del estudiante y el Número de cursos repetidos —un indicador clásico de dificultades académicas y predictor del abandono—, junto con la clasificación de la institución (por ejemplo, Institución Educativa Indígena), la Zona (rural o urbana), la condición laboral del padre, los niveles de estudio de la madre o tutor y el distrito del local escolar. Aunque estas variables se seleccionan con menor consistencia, complementan el modelo al capturar aspectos indirectos pero relevantes para la permanencia escolar. Por ejemplo, en esta cohorte, los hijos de madres que concluyeron la educación

primaria registran una tasa de abandono cercana al 13%, que disminuye al 8% si la madre completó la educación secundaria y alcanza sólo el 4% en el caso de madres con estudios universitarios finalizados. Del mismo modo, aquellos estudiantes cuyos padres desempeñan ocupaciones laborales estables (empleados públicos o privados, empleadores) exhiben tasas de abandono de entre 7% y 10%, mientras que en ocupaciones más vulnerables, la tasa de abandono puede elevarse hasta 15%–19%. Estos resultados concuerdan con la literatura que resalta la influencia de las condiciones socioeconómicas en la trayectoria educativa. La menor aparición de estas variables en algunos métodos de selección podría deberse al solapamiento con otros predictores que recogen dimensiones similares, de modo que, aunque el nivel de estudio de la madre o la condición laboral del padre no figuren explícitamente en ciertos modelos, sus efectos pueden quedar reflejados a través de variables correlacionadas.

Modelos Predictivos del Abandono Escolar

En la Tabla VIII se resumen las métricas de desempeño obtenidas por cada uno de los modelos implementados para la predicción del abandono escolar. Se incluyen: exactitud (Accuracy), precisión (Precision), sensibilidad (Recall), especificidad (Specificity) y F1-Score.

TABLA VIII. Rendimiento de Modelos Predictivos para la Predicción del Abandono Escolar.

Modelo	Accuracy	Precision	Recall	Specificity	F1-Score
Regresión Logística	0.7332	0.2436	0.5980	0.7513	0.3462
Random Forest	0.7054	0.2269	0.6207	0.7167	0.3323
LASSO	0.7134	0.9312	0.7288	0.5980	0.8177
XGBoost	0.7062	0.2283	0.6247	0.7172	0.3344
SVM	0.7022	0.2227	0.6107	0.7145	0.3264
Redes Neuronales	0.7198	0.2364	0.6154	0.7338	0.3416

Fuente: Base de datos RUE, elaboración propia.

Considerando los resultados de la **TABLA VIII**, la Regresión Logís-

tica presenta la mayor exactitud (0.7332), lo que implica que el 73.32% de las predicciones (abandono y no abandono) son correctas, y su *Recall* (0.5980) indica que detecta cerca del 60% de los estudiantes que abandonan. Sin embargo, su precisión es baja (0.2436), lo que significa que una proporción significativa de las predicciones positivas corresponde a casos de no abandono.

Random Forest y XGBoost exhiben desempeños similares, con valores de *Recall* superiores a 0.62, lo que beneficia la identificación de estudiantes en riesgo; no obstante, su precisión es limitada (≈ 0.22), reflejándose en F1-Scores relativamente bajos (≈ 0.33). SVM también muestra un patrón similar, con *Recall* de 0.6107 y precisión de 0.2227, además de un Accuracy y F1-Score comparativamente menores. Las Redes Neuronales obtienen una exactitud intermedia (0.7198) y un *Recall* de 0.6154, valores comparables a los de la Regresión Logística.

La Regresión LASSO se destaca por su alta precisión (0.9312) y *Recall* (0.7288), lo que se traduce en un F1-Score de 0.8177, el más elevado entre todos los modelos evaluados. Aunque su exactitud es de 0.7134, su sensibilidad la convierte en una opción sólida para identificar a los estudiantes en riesgo, mientras que su elevada precisión ayuda a minimizar las falsas alarmas. Con base en los valores de la Tabla VIII, se puede concluir que LASSO ofrece el mejor rendimiento global en la predicción del abandono escolar a partir de las variables seleccionadas.

Además de su sólido desempeño predictivo, la Regresión LASSO proporciona un modelo parsimonioso al descartar coeficientes sin efecto significativo, manteniendo la interpretabilidad propia de la Regresión Logística. Por ejemplo, para la variable Edad, el coeficiente estimado ($\beta \approx 0.847$) se traduce en un Odds Ratio de $\exp(0.847) \approx 2.33$, lo que significa que, manteniendo constantes las demás variables, un estudiante un año mayor tiene 2.33 veces mayor probabilidad de abandonar.

La interpretación de las demás variables es menos directa, ya que se trata de predictores categóricos convertidos en variables ficticias (dummies). Sin embargo, el OR permite evaluar intuitivamente su impacto: la variable Turno es la segunda más influyente ($OR \approx 1.13$), mientras que Departamento ($OR \approx 1.02$) y Especialidad ($OR \approx 0.99$) presentan un efecto más limitado. Además, el coeficiente asociado a Sobreedad se reduce a cero ($OR = 1$), lo que indica que, en este modelo penalizado, la sobreedad no añade un impacto significativo, probablemente porque su efecto ya está capturado por la variable Edad.

En el contexto de este estudio, la Regresión LASSO exhibe un des-

empeño particularmente robusto al conjugar alta sensibilidad y precisión. Su capacidad de regularización mediante penalización L1 facilita la eliminación de variables con menor relevancia, dando lugar a un modelo parsimonioso e interpretable que prioriza los predictores con mayor incidencia en el abandono. Bajo estas condiciones, LASSO se consolida como la alternativa más adecuada para la identificación temprana de estudiantes en riesgo, al minimizar tanto los falsos positivos como la complejidad del modelo.

Conclusiones

La presente investigación se fundamentó en una extensa base de datos del Registro Único del Estudiante (RUE) del Ministerio de Educación y Ciencias (MEC) de Paraguay, abarcando el periodo 2017–2023 y un elevado número de registros. Las fortalezas metodológicas, en particular el uso de datos oficiales y técnicas analíticas rigurosas, aportan evidencia sólida sobre el abandono escolar en la Educación Media en Paraguay. Además, el reducido número de estudios en América Latina, basados en datos administrativos oficiales y centrados en niveles preuniversitarios, refuerza el aporte de este trabajo.

El análisis descriptivo revela un incremento sostenido en la matrícula entre 2017 y 2021, con picos de abandono en 2019 y 2020. La irrupción de la pandemia en 2020 agravó la deserción en un contexto en el que la educación adoptó la modalidad a distancia; en 2021, los abandonos disminuyeron en un escenario de implementación de medidas de contención y retorno parcial a la presencialidad. En cuanto al género, aunque inicialmente la tasa de abandono en hombres era superior, a partir de 2019 se observa un aumento progresivo en la proporción femenina. Además, la edad promedio de los estudiantes aumentó de 2017 a 2023, lo que indica un incremento en la sobreedad o retraso escolar.

Geográficamente, los patrones de abandono son similares en zonas urbanas y rurales; sin embargo, la tasa es significativamente más alta en las instituciones públicas en comparación con las privadas y/o subvencionadas. Se observan diferencias notables entre los Departamentos, y el turno Noche registra la tasa más elevada en comparación con otros turnos. En cuanto a la especialidad, las modalidades “Sin Especialidad” y “Formación Profesional Vocacional” presentan los porcentajes más altos. Además, la primera repetición de curso se evidencia como un factor crítico y desencadenante del abandono.

La selección de atributos destacó la relevancia de la edad (especialmente la sobreedad), el tipo de especialidad, el turno y el Departamento como variables que explican de forma consistente la deserción. Una edad avanzada dentro del mismo curso aumenta notablemente el riesgo de abandono. Asimismo, los Departamentos reflejan desigualdades territoriales, y el turno Noche en Paraguay se asocia a desafíos laborales y familiares. Además, especialidades como la Formación Profesional Vocacional reúnen a estudiantes con necesidades inmediatas de inserción laboral o con menores perspectivas de continuar estudios, lo que se traduce en tasas de abandono elevadas. Si bien la sobreedad y la repetición ya se habían reconocido como predictores sólidos (Conde et al., 2023), nuestro estudio amplía el espectro al evidenciar con especial nitidez la incidencia de las modalidades de bachillerato y del turno Noche, factores que hasta ahora habían recibido escasa atención en la literatura.

Las comparaciones de desempeño en los Modelos Predictivos muestran que la Regresión LASSO equilibra sensibilidad y precisión de manera sobresaliente, identificando cerca del 73% de los casos de abandono (Recall) y acertando en el 93% de las predicciones positivas (Precision). Este balance se refleja en un F1-Score de ≈ 0.82 , el más alto entre todos los modelos, lo que la convierte en la técnica más prometedora para la identificación temprana de estudiantes en riesgo. Su robustez e interpretabilidad la hacen ideal para orientar políticas de alerta temprana, detectando eficazmente a los alumnos con mayor probabilidad de abandonar y minimizando falsas alarmas.

Las implicaciones prácticas del estudio muestran la necesidad de diseñar intervenciones que aborden los factores identificados en el análisis, tanto estructurales (desigualdades entre departamentos y apoyo específico a los turnos nocturnos) como individuales (cursos repetidos, sobreedad y orientación vocacional). Se recomienda que los responsables de la política educativa refuercen sus acciones en las instituciones oficiales, donde se registran las tasas más altas de abandono, sobre todo en el turno noche y en las modalidades “Sin Especialidad” y “Formación Profesional Vocacional”, y prestando particular atención a los estudiantes con al menos un curso repetido. Además, es fundamental elaborar planes de orientación y tutoría que incluyan a las familias, de modo que las acciones se coordinen dentro y fuera del ámbito escolar. Conforme a la propuesta de Gutiérrez-de-Rozas et al. (2023), estos planes deberían profundizar en dimensiones no cognitivas (autoestima, habilidades sociales y de vida, toma de decisiones) y en estrategias de enseñanza-aprendizaje.

Como contribución a la literatura especializada, este estudio demuestra la utilidad de los registros administrativos oficiales de gran escala para el análisis del abandono escolar en América Latina. A través de la vinculación de características sociodemográficas, académicas e institucionales de más de 700,000 estudiantes, se confirma que la sobreedad y la trayectoria escolar previa (en particular las repeticiones) siguen siendo predictores robustos de este fenómeno. Asimismo, se constata que ciertas modalidades de bachillerato, la asistencia en el turno nocturno y las brechas territoriales reflejan cómo las desigualdades estructurales siguen incidiendo de manera significativa en el fenómeno del abandono escolar. Este trabajo no solo amplía la evidencia internacional sobre la influencia de factores socioeconómicos y escolares en la desvinculación educativa (Hernández-Leal et al., 2021), sino que también pone de relieve el potencial analítico de los datos censales para informar políticas de retención, especialmente en contextos donde los Sistemas de Información y Gestión Educativa aún se encuentran en fases incipientes (Arias et al., 2021).

Entre las limitaciones del estudio se destaca la cobertura y calidad de algunos campos del RUE, y la inclusión de factores psicosociales y económicos podría mejorar la precisión de los modelos analíticos. Como líneas de investigación futura, se sugiere ampliar el rango temporal para evaluar mejor los efectos de la pandemia en el abandono escolar en Paraguay, profundizar en variables cualitativas (motivación, clima escolar, rol familiar) y explorar metodologías alternativas, tales como modelos de series temporales y enfoques de análisis causal, integrando además bases de datos externas (programas sociales, estadísticas de salud) para obtener una visión más integral del riesgo de abandono. Adicionalmente, resultaría valioso comparar los resultados obtenidos en este trabajo con metodologías emergentes de inteligencia artificial, como por ejemplo redes neuronales profundas, redes recurrentes y arquitecturas basadas en transformadores, capaces de capturar patrones secuenciales y complejos en la trayectoria educativa (Kusumawardani y Alfa-rozi, 2023; Nguyen et al., 2024).

En suma, este trabajo contribuye a la comprensión del abandono escolar en la educación secundaria de Paraguay y de América Latina, destacando la utilidad de la analítica de datos para la formulación de políticas. La adopción de estas recomendaciones y la mejora continua de los sistemas de información podrían fortalecer la capacidad de respuesta de las autoridades educativas, impactando positivamente en la permanencia y graduación de los estudiantes.

Referencias bibliográficas

- Abideen, Z. U., Mazhar, T., Razzaq, A., Haq, I., Ullah, I., Alasmay, H., & Mohamed, H. G. (2023). Analysis of Enrollment Criteria in Secondary Schools Using Machine Learning and Data Mining Approach. *Electronics*, 12(3), 694. <https://doi.org/10.3390/electronics12030694>
- Ajjawi, R., Dracup, M., Zacharias, N., Bennett, S., & Boud, D. (2020). Persisting students' explanations of and emotional responses to academic failure. *Higher Education Research & Development*, 39(2), 185-199. <https://doi.org/10.1080/07294360.2019.1664999>
- Alladatin, J., Lionel, R., & Insaf, A. C. (2023). School principal's training programs, challenges, and improvement opportunities: rapid review. *International Journal of Educational Innovation and Research*, 3(1), 17-26. <https://doi.org/10.31949/ijeir.v3i1.6830>
- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Amaya-Amaya, A., Huerta-Castro, F., & Flores-Rodríguez, C. O. (2020). Big Data, una estrategia para evitar la deserción escolar en las IES. *Revista Iberoamericana de Educación Superior*, 11(31), 166-178. <https://doi.org/10.22201/iisue.20072872e.2020.31.712>
- Arias Ortiz, E., Eusebio, J., Pérez Alfaro, M., Vásquez, M. y Zoido, P. (2021). *Los Sistemas de Información y Gestión Educativa (SIGED) de América Latina y el Caribe: la ruta hacia la transformación digital de la gestión educativa*. <https://doi.org/10.18235/0003345>
- Conde Vélez, S., García Rodríguez, M. D. P., & Toscano Cruz, M. D. L. O. (2023). Riesgo de abandono escolar: ¿cómo influyen las características sociofamiliares percibidas por los estudiantes sobre sus actitudes y comportamiento en el aula?. *Educación XXI*, 26(2), 267-298. <https://doi.org/10.5944/educxx1.33279>
- Comisión Económica para América Latina y el Caribe (enero de 2024). *Prevención y reducción del abandono escolar en América Latina y el Caribe*. Naciones Unidas. <https://www.cepal.org/es/publicaciones/68814-prevencion-reduccion-abandono-escolar-america-latina-caribe>

- European Education Area (17 de junio de 2022). *Early School Leaving*. European Commision <https://education.ec.europa.eu/es/education-levels/school-education/early-school-leaving>.
- EUROSTAT (19 de mayo de 2025). *Early leavers from education and training*. European Union [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Early leavers from education and training](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Early_leavers_from_education_and_training)
- Flores Satalaya, J. M. (2025). El machine learning para abordar el abandono escolar: Una revisión de los modelos más innovadores. *Ciencia Latina Revista Científica Multidisciplinar*, 8(6), 10993-11027. https://doi.org/10.37811/cl_rcm.v8i6.15824
- Gutiérrez-de-Rozas, B., López-Martín, E., & Carpintero, E. (2023). Análisis del abandono educativo temprano en España: condicionantes del abandono y motivos para regresar al sistema educativo. *Revista de Investigación Educativa*, 41(2), 523-549. <https://doi.org/10.6018/rie.546521>
- Hernández-Leal, E., Duque-Méndez, N. D., & Cechinel, C. (2021). Unveiling educational patterns at a regional level in Colombia: data from elementary and public high school institutions. *Heliyon*, 7(9), e08017. <https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e08017>
- Hernández Prados, M.A., & Alcaraz Rodríguez, M. (2018). Factores incidentes en el abandono escolar prematuro. *Revista de Investigación en Educación*, 16(2), 182-195. <https://revistas.uvigo.es/index.php/reined/article/view/2136/2164>
- Insfrán-Coronel, D.R., Enrique-Sánchez, E.M., Beck, F., López-Fernández, A., & García-Torres, M. (2024). Analysis of School Dropout Rate in Paraguay Using a Machine Learning Approach en Quintián, H., et al. (Eds.), *International Joint Conferences*. (Vol. 957). Springer. https://doi.org/10.1007/978-3-031-75016-8_29
- Ministerio de Educación y Ciencias (2013). La deserción escolar en Paraguay: Características que asume en la educación media <https://desarrollo.org.py/admin/app/webroot/pdf/publicaciones/08-10-2015-16-45-20-848785920.pdf>
- Instituto Nacional de Estadística (26 de febrero de 2024). *El INE comparte datos sobre la educación paraguaya* <https://www.ine.gov.py/noticias/1890/el-ine-comparte-datos-sobre-la-educacion-paraguaya>

- Khurram, F., Hossain, M., Korobi, F., Khan, M., Talukder, M., Nath, C., & Talukder, A. (2023) Understanding School Dropout and Its Impact on the Community in the Khulna Region of Bangladesh. *Open Journal of Social Sciences*, 11(11), 582-596. <https://doi.org/10.4236/ojss.2023.1111038>
- Kocsis, Á., & Molnár, G. (2024). Factors influencing academic performance and dropout rates in higher education. *Oxford Review of Education*, 51(3), 414–432. <https://doi.org/10.1080/03054985.2024.2316616>
- Krüger, J. G., Britto Jr., A. de S., & Barddal, J. P. (2023). An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications*, 233, 120933. <https://doi.org/10.1016/j.eswa.2023.120933>
- Kusumawardani, S. S., & Alfarozi, S. A. I. (2023). Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities. *IEEE Access*, 11, 18960–18971. <https://doi.org/10.1109/access.2023.3246122>
- Ministerio de Educación, Formación Profesional y Deportes (7 de febrero de 2025). *Nivel de formación, Formación permanente y Abandono: Explotación de las variables educativas de la Encuesta de Población Activa*
<https://www.educacionfpydeportes.gob.es/servicios-al-ciudadano/estadisticas/laborales/epa.html>
- Montero-Sieburth, M., & Turcatti, D. (2022). Preventing disengagement leading to early school leaving: pro-active practices for schools, teachers and families. *Intercultural Education*, 33(2), 139-155. <https://doi.org/10.1080/14675986.2021.2018404>
- Nait Belaid, Y. (2021). A critical analysis of policies to fight illiteracy and school dropout in rural Morocco. *International Journal of Advanced Research*, 9 (08), 254-269. <https://dx.doi.org/10.21474/IJAR01/13259>
- Nguyen Thi Cam, H., Sarlan, A., & Arshad, N. I. (2024). A hybrid model integrating recurrent neural networks and the semi-supervised support vector machine for identification of early student dropout risk. *PeerJ Computer Science*, 10, e2572. <https://doi.org/10.7717/peerj-cs.2572>
- Ortiz-Lozano, J. M., Aparicio-Chueca, P., Triadó-Ivern, X. M., & Arroyo-Barrigüete, J. L. (2023). Early dropout predictors in social scienc-

- es and management degree students. *Studies in Higher Education*, 49(8), 1303–1316. <https://doi.org/10.1080/03075079.2023.2264343>
- Ramsdal, G. H., & Wynn, R. (2022). Theoretical Basis for a Group Intervention Aimed at Preventing High School Dropout: The Case of ‘Guttas Campus’. *International Journal of Environmental Research and Public Health*, 19(24), 17025. <https://doi.org/10.3390/ijerph192417025>
- Rodríguez, P., Delgado Villanueva, A., Dombrowskaia, L., & Valenzuela, J. P. (2023). A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: the case of Chile. *Education and Information Technologies*, 28(8), 10103–10149. <https://doi.org/10.1007/s10639-022-11515-5>
- Segura, M., Mello, J., & Hernández, A. (2022). Machine learning prediction of university student dropout: Does preference play a key role?. *Mathematics*, 10(18), 3359. <https://doi.org/10.3390/math10183359>
- Serrano, K., Sierra, Y., Rios-Cruz, S., & Pardo, F. (2024). Machine Learning models for predicting academic withdrawal in college students. En *Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, 1-7. [10.1109/CONIITI64189.2024.10854829](https://doi.org/10.1109/CONIITI64189.2024.10854829)
- Smith Uldall, J., & Gutiérrez Rojas, C. (2022). An Application of Machine Learning in Public Policy: Early Warning Prediction of School Dropout in the Chilean Public Education System. *Multidisciplinary Business Review*, 15(1), 20–35. <https://doi.org/10.35692/07183992.15.1.4>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (2022). *La encrucijada de la educación en América Latina y el Caribe: Informe regional de monitoreo ODS4–Educación 2030*. UNESCO Oficina Regional de Educación para América Latina y el Caribe. https://unesdoc.unesco.org/ark:/48223/pf0000382919_spa
- Venkatesan R.G., & Mappillairaju, B. (2023). Detection of hotspots of school dropouts in India: A spatial clustering approach. *PLoS ONE*, 18(1), e0280034. <https://doi.org/10.1371/journal.pone.0280034>
- Zapata-Medina, D., Espinosa-Bedoya, A., & Jiménez-Builes, J. A. (2024). Improving the Automatic Detection of Dropout Risk in Middle and High School Students: A Comparative Study of Feature Selection Techniques. *Mathematics*, 12(12), 1776. <https://doi.org/10.3390/math12121776>

Zengin, M. (2021). Investigation of High School students' dropout risk level. *Shanlax International Journal of Education*, 9, 59-68. <https://doi.org/10.34293/education.v9is1-may.4000>

Información de contacto: Jorge Daniel Mello-Román. Universidad Nacional de Concepción. E-mail: jdmello@facet-unc.edu.py

