


Sistema de Análisis de Datos Basado en Machine Learning para Identificar Patrones entre Deserción Escolar y Pobreza Multidimensional en Paraguay en el Contexto de la Informática Empresarial.

Machine Learning-Based Data Analysis System to Identify Patterns between School Dropout and Multidimensional Poverty in Paraguay in the Context of Business Informatics.

Insfrán Coronel, Diego Rubén ¹ 

Sánchez, Enzo Marcel Enrique ² 

Navarro Garay; Angela ³ 

Autor de correspondencia: Diego Rubén Insfrán Coronel, email: diegoinsfran8@gmail.com

Como referenciar este artículo

Insfrán, D., Sánchez, E. y Navarro, A. (2025). Sistema de Análisis de Datos Basado en Machine Learning para Identificar Patrones entre Deserción Escolar y Pobreza Multidimensional en Paraguay en el Contexto de la Informática Empresarial. Revista Científica Multidisciplinaria Tajy. 2025; 2(1): 73-85

Resumen

Este estudio investiga la relación entre la tasa de deserción escolar y el índice de pobreza multidimensional en Paraguay mediante un proceso de extracción, transformación y carga (ETL), se analizaron datos del Ministerio de Educación y Ciencias (MEC) y del Instituto Nacional de Estadística (INE). La investigación emplea técnicas de agrupamiento, en particular el método K-Means, para identificar patrones y perfiles de riesgo entre los estudiantes. Los hallazgos destacan el impacto significativo de factores socioeconómicos, como la pobreza y el trabajo infantil, en las tasas de deserción escolar. Estos hallazgos buscan fundamentar intervenciones específicas para mejorar los resultados educativos y reducir las tasas de deserción escolar en Paraguay.

Palabras clave: educación, deserción escolar, agrupamiento, minería de datos.

Abstract

This study investigates the relationship between the school dropout rate and the multidimensional poverty index in Paraguay through an extraction, transformation, and loading (ETL) process, data from the Ministerio de Educación y Ciencias (MEC) and the Instituto Nacional de Estadísticas (INE) were analyzed. The research employs clustering techniques, particularly the K-Means method, to identify patterns and risk profiles among students.

¹ Universidad Nacional de Concepción, Facultad de Ciencias Económicas y Administrativas, Docente, diegoinsfran8@gmail.com

² Universidad Nacional de Concepción, Facultad de Ciencias Económicas y Administrativas, Docente, nzosanz@gmail.com

³ Universidad Nacional de Concepción, Facultad de Ciencias Económicas y Administrativas, Docente, anavarro@fcea-unc.edu.py



Findings highlight the significant impact of socioeconomic factors, such as poverty and child labor, on school dropout rates. These results aim to inform targeted interventions to improve educational outcomes and reduce school dropout rates in Paraguay.

Keywords: education, school dropout , clustering, data mining, school dropout.

Introducción

En el Paraguay, cada año, una considerable cantidad de adolescentes abandona sus estudios al pasar del noveno grado al primer curso de la educación escolar media. Este fenómeno, que se observa principalmente en instituciones públicas, no solo implica una interrupción del proceso educativo, sino que también acarrea consecuencias negativas a nivel social, económico y cultural para el país. La deserción escolar, influenciada por factores familiares, económicos y pedagógicos, se ha convertido en una problemática estructural que requiere atención urgente desde múltiples enfoques (Reyes, G.E., Sosa, D.V., Quispe, M.G., Iraola-Real, I., 2021).

Para comprender este fenómeno de manera integral, en este trabajo se plantea el uso de técnicas de minería de datos y aprendizaje automático (machine learning), herramientas fundamentales en la actualidad para el análisis de grandes volúmenes de datos. A través de estas técnicas es posible identificar patrones, correlaciones y factores ocultos que inciden en la decisión de

abandonar los estudios. El uso del proceso ETL (extracción, transformación y carga) garantiza la integridad y confiabilidad de los datos provenientes de fuentes gubernamentales, tales como el Ministerio de Educación y Ciencias (MEC) y el Instituto Nacional de Estadística (INE).

El análisis se centra en estudiantes de entre 10 y 17 años de edad que cursan en instituciones públicas del país, con énfasis en aquellos que transitaron del noveno grado al primer año de la educación escolar media. La aplicación del algoritmo de agrupamiento K-Means permite segmentar a los estudiantes en función del riesgo de deserción, lo cual contribuye a entender de manera más precisa las causas que impulsan el abandono escolar.

Los modelos de agrupamiento utilizados proporcionan información fundamental para la formulación de políticas públicas eficaces contra la deserción escolar. Con este fin, el estudio busca no solo identificar a los estudiantes en situación de riesgo, sino también ofrecer datos accionables para implementar intervenciones tempranas y basadas en evidencia.

Descripción del problema

La deserción escolar en Paraguay es un problema persistente y profundamente arraigado en los sectores más vulnerables de la sociedad. Las precarias condiciones socioeconómicas, en particular la situación económica del hogar, inciden significativamente en la continuidad de los estudiantes en el sistema educativo. A medida que estas condiciones se vuelven más desfavorables, disminuyen las oportunidades de que los jóvenes permanezcan en las aulas y adquieran las competencias necesarias para insertarse dignamente en el mercado laboral (Reyes, G.E., Sosa, D.V., Quispe, M.G., Iraola-Real, I., 2021).

Entre los factores más relevantes que inciden en esta problemática se encuentra la pobreza, la cual limita la capacidad de las familias para sostener la escolarización de sus hijos y la del propio sistema educativo para brindar un acompañamiento adecuado. Si bien existen esfuerzos estatales para mitigar esta situación estas medidas aún no han logrado un impacto sostenido (Gobiernodelparaguay.gov.py, 2024).

Este fenómeno se agrava durante la continuidad en la etapa educativa de muchos adolescentes que abandonan sus estudios para integrarse al mercado laboral, motivados por la necesidad económica o por la percepción de que la educación ya no les representa un beneficio inmediato. En este punto crítico del recorrido académico, la combinación de factores sociales,

económicos y pedagógicos impulsa a muchos jóvenes fuera del sistema educativo formal, truncando sus posibilidades de desarrollo personal y profesional a largo plazo (Reyes, G.E., Sosa, D.V., Quispe, M.G., Iraola-Real, I., 2021).

La comprensión del comportamiento de la deserción escolar en este tramo educativo, así como la identificación de los factores que la condicionan, resulta esencial para orientar estrategias de política pública eficaces y sostenibles. Por ello, el presente estudio se plantea como una vía para indagar sistemáticamente cómo se manifiesta este fenómeno en el país, específicamente en el año 2022, prestando atención a las variables socioeconómicas y académicas involucradas.

Objetivo

El objetivo de esta investigación es analizar la relación entre la tasa de deserción escolar y el Índice de Pobreza Multidimensional en el tránsito del noveno grado al primer año de la educación media en el Paraguay durante el año 2022.

Reseña histórica

La deserción escolar en Paraguay es un problema persistente, especialmente en la crucial transición del noveno grado al primer año de la educación media. Cada año, una cantidad considerable de adolescentes interrumpe su trayectoria educativa, lo que genera consecuencias negativas a nivel social, económico y cultural para el país (Reyes, G.E., Sosa, D.V., Quispe, M.G.,

Iraola-Real, I., 2021). Si bien se reconoce que factores familiares, económicos y pedagógicos influyen en este fenómeno, existe una laguna significativa en la comprensión de los patrones y las relaciones específicas entre estos factores y la deserción (Gobiernodelparaguay.gov.py, 2024). Actualmente, no se dispone de una base de datos unificada o un análisis sistemático que vincule de forma directa las causas subyacentes con los casos de abandono.

En lo teórico, a diferencia de los enfoques tradicionales, este estudio propone una metodología innovadora que integra técnicas avanzadas de minería de datos y aprendizaje automático. Esto permite ir más allá de la estadística descriptiva para descubrir patrones ocultos, correlaciones y factores influyentes que no son evidentes a primera vista. Se utilizaron conjuntos de datos que, si bien son diferentes en su origen, son procesados y analizados de forma complementaria. Un riguroso proceso de ETL garantiza la integridad y confiabilidad de estos datos, crucial para obtener resultados precisos.

En lo metodológico, la aplicación del algoritmo de agrupamiento K-Means es fundamental. Esta técnica permite segmentar a los estudiantes en distintos grupos de riesgo de deserción, ofreciendo una visión detallada de los perfiles más vulnerables. Por ejemplo, al combinar datos de inscripción con el índice de pobreza multidimensional, podemos

identificar la relación directa entre las condiciones socioeconómicas y la probabilidad de abandono. Esta capacidad de perfilar el riesgo de deserción en estudiantes de escuelas públicas de 10 a 17 años es un aporte innovador y muy práctico que establece un nuevo camino para análisis futuros.

Y finalmente en lo práctico, los hallazgos de esta investigación son de gran utilidad y relevancia social. En primer lugar, generan información clave que sirve como base sólida para la formulación de políticas públicas más efectivas y focalizadas. Al conocer los factores de riesgo y los grupos más afectados, las autoridades educativas pueden diseñar intervenciones tempranas y específicas, optimizando los recursos y garantizando la igualdad de oportunidades. En segundo lugar, esta metodología establece un precedente para futuras investigaciones en el ámbito educativo de Paraguay, demostrando el potencial de la analítica de datos. Los resultados no solo proporcionan una ruta clara para replicar y mejorar estos análisis, sino que también abren la puerta a la exploración de algoritmos más avanzados o la integración de nuevas fuentes de datos para un entendimiento aún más profundo del fenómeno.

El índice de pobreza multidimensional

El Índice de Pobreza Multidimensional (IPM) es una métrica crucial que identifica múltiples privaciones que afectan a hogares e

individuos en dimensiones esenciales como la salud, la educación y el nivel de vida (Alkire & Foster, Counting and Multidimensional Poverty Measurement, 2008), (World Bank Group, s.f.). A diferencia de los enfoques tradicionales que se centran únicamente en el ingreso, el IPM ofrece una visión mucho más integral y holística de las carencias, lo que lo posiciona como una herramienta fundamental para una comprensión profunda de la pobreza (Alkire & Foster, Counting and Multidimensional Poverty Measurement, 2008).

La elección de utilizar el IPM en esta investigación se fundamenta en su superior capacidad para capturar la complejidad inherente de la pobreza y su impacto en fenómenos sociales como la deserción escolar a nivel nacional en Paraguay. Los indicadores monetarios, si bien útiles para medir la privación de ingresos, resultan insuficientes para explicar las múltiples facetas que influyen en la decisión o necesidad de un estudiante de abandonar sus estudios. La deserción rara vez se explica solo por la falta de recursos económicos; a menudo, es el resultado de una acumulación de carencias en diversas dimensiones de la vida de un hogar. Un hogar puede, por ejemplo, superar la línea de pobreza monetaria, pero aun así carecer de acceso a una educación de calidad, a servicios de salud adecuados o a condiciones de vivienda dignas, todos ellos factores que, directa o

indirectamente, pueden empujar a un estudiante fuera del sistema educativo (Alkire & Santos, Acute Multidimensional Poverty: A New Index for Developing Countries, 2010).

En este sentido, el IPM, al considerar privaciones simultáneas en salud, educación y nivel de vida, proporciona una medida más pertinente y robusta para el análisis de la deserción escolar. Su metodología permite identificar a aquellos individuos y hogares que sufren carencias entrelazadas, revelando patrones de vulnerabilidad que un análisis basado solo en ingresos no podría discernir. Por ejemplo, un estudiante podría abandonar la escuela no únicamente por dificultades económicas, sino también debido a una enfermedad en la familia (dimensión de salud), la necesidad de trabajar por la falta de empleo de los padres (dimensión de nivel de vida/empleo) o la lejanía de una escuela accesible (dimensión educativa) (Alkire & Santos, Acute Multidimensional Poverty: A New Index for Developing Countries, 2010).

En el contexto paraguayo, el Instituto Nacional de Estadística (INE) calcula y publica regularmente el IPM (Instituto Nacional de Estadística (INE), 2022), lo cual garantiza la disponibilidad de datos confiables y asegura la coherencia de esta investigación con los marcos de política pública nacionales de monitoreo de la pobreza. Esta alineación fortalece la validez y la aplicabilidad de los hallazgos de este

estudio, permitiendo una comprensión más profunda de cómo las múltiples privaciones socioeconómicas actúan como un motor de la deserción escolar en el país. Para esta investigación, el IPM y sus dimensiones serán variables clave para caracterizar el contexto socioeconómico de los jóvenes y su relación con la deserción escolar, especialmente en la crítica transición del noveno grado al primer curso de la educación media.

Materiales y Métodos

Esta investigación adoptó un enfoque cuantitativo, de tipo correlacional con alcance explicativo y se desarrolló bajo un diseño no experimental de corte transversal (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014) (Sabino, 2017). Se trabajó con datos abiertos del Ministerio de Educación y Ciencias (MEC) y del Instituto Nacional de Estadística (INE) del Paraguay, correspondientes al año 2022. Los datos fueron tratados como fuentes secundarias y sirvieron para analizar la deserción escolar en la transición del noveno grado al primero de la educación media (Sabino, 2017).

Los datos fueron integrados mediante un proceso ETL (Extracción, Transformación y Carga), que permitió unificar, limpiar y estructurar la información adecuadamente (Inmon, 2005) (Kimball & Ross, 2013). Posteriormente, se aplicaron técnicas de minería de datos y aprendizaje automático no

supervisado, específicamente el algoritmo K-Means (Jain, A. K., 2010), con el objetivo de agrupar a los estudiantes según características similares y detectar perfiles de riesgo de deserción. Para determinar el número óptimo de clústeres, se utilizó el método de la silueta (Rousseeuw, 1987), evaluando su puntaje con una muestra de 50.000 registros.

En cuanto al procesamiento y análisis de datos, se utilizó el lenguaje de programación Python en el entorno PyCharm Community (JetBrains, 2025), empleando librerías especializadas como Pandas, NumPy, Scikit-learn, Matplotlib y Seaborn (McKinney, 2010), (NumPy Developers, 2023), (Pedregosa, y otros, 2011), (Matplotlib Development Team, 2023), (Seaborn Developers, 2025). Estas herramientas facilitaron la transformación de variables, la normalización, la codificación, la implementación del modelo K-Means y la visualización de resultados mediante gráficos descriptivos y de agrupamiento.

Resultados

El pipeline genera resúmenes estadísticos clave como medias, desviaciones estándar y porcentajes de incidencia de trabajo infantil por clúster, lo que permite obtener una visión clara de las distribuciones y variaciones dentro de cada grupo. Además, calcula la brecha entre áreas rurales y urbanas, facilitando la identificación de disparidades.

Esta información es crucial para comprender la relación entre variables y definir áreas de intervención específica. La capacidad de guardar los resultados en formato CSV asegura que los hallazgos puedan ser utilizados en análisis posteriores o integrados con otros sistemas de información.

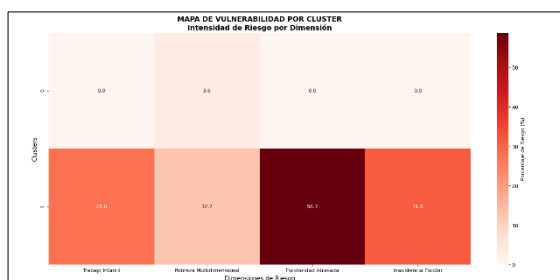


Figura N° 1. Mapa de vulnerabilidad por clúster.

La investigación, utilizando el algoritmo K-Means y validando los resultados con el método de la silueta, concluyó que el número ideal de agrupaciones o clústeres era dos ($K=2$), mediante la técnica de muestreo estratégico. A lo largo del análisis, estos clústeres se han denominado consistentemente clúster 0 y clúster 1, permitiendo interpretar las características de los estudiantes dentro de ellos y describirlas como tres perfiles diferenciados.

El primer perfil, de alta vulnerabilidad social y económica, está directamente relacionada con el clúster 1 y se distingue por la acumulación de factores de riesgo severos. Los estudiantes que pertenecen a este grupo provienen de hogares donde la pobreza multidimensional es alarmante (33% “Pobre No Extremo” y 14% “Pobre Extremo”). Una de las preocupaciones más serias es la

notable presencia de trabajo infantil entre los jóvenes de 10 a 17 años, que se encuentra en el 29% de los hogares, y que se eleva al 46% en las áreas rurales de este clúster. Además, este perfil revela un alto índice de escolaridad atrasada (59% de los hogares) y una preocupante tasa de inasistencia escolar (32% de los hogares), especialmente en las zonas rurales, donde las limitaciones de recursos y las presiones laborales son mucho más intensas.

Un segundo perfil, de vulnerabilidades moderadas, también se desprende del clúster 1, pero representa una subcaracterización de este grupo con condiciones menos extremas. Este perfil se distingue por presentar índices de pobreza intermedios y una menor proporción de menores en situación laboral en comparación con el perfil de alta vulnerabilidad. Se sugiere que estos estudiantes podrían encontrarse en una transición entre zonas rurales y urbanas, donde la oferta de modalidades educativas podría ser más limitada.

Un tercer perfil, con ventajas socioeconómicas relativas, se corresponde con el clúster 0 y se caracteriza por menores índices de pobreza (81% de hogares “No Pobre”), ausencia de trabajo infantil (100% de hogares “No trabaja”), y una asistencia escolar total con no presencia de escolaridad atrasada. La identificación de este perfil es crucial, ya que sugiere que la deserción escolar, aunque en menor medida, no se

limita únicamente a la pobreza extrema, sino que puede manifestarse de diversas maneras, incluso en contextos con mejores condiciones socioeconómicas.

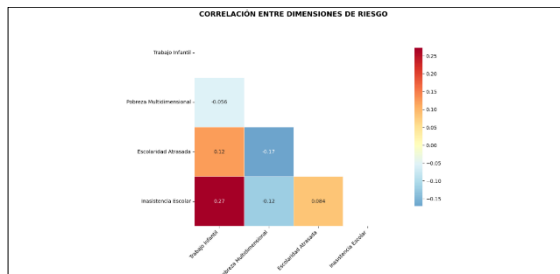


Figura N° 2. Correlación entre dimensiones de riesgo.

La diferenciación por modalidades educativas proporciona a través de los clústeres información valiosa sobre como diferentes opciones formativas atraen o retienen a estudiantes específicos. La educación técnica muestra patrones de matriculación que sugieren mayor atracción para estudiantes de familias con condiciones socioeconómicas intermedias, posiblemente porque ofrece perspectivas de inserción laboral más inmediatas que la educación científica. La educación científica, por su parte, presenta concentraciones mayores en estudiantes de sectores urbanos y con menores índices de pobreza, sugiriendo que el acceso a esta modalidad puede estar condicionado por factores socioeconómicos y geográficos específicos.

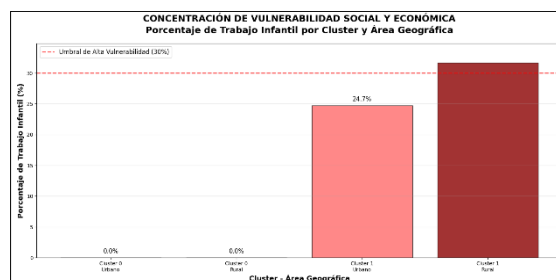


Figura N° 3. Concentración de vulnerabilidad social y económica.

La identificación de estos perfiles diferenciados tiene implicaciones directas para el diseño de políticas educativas específicas. Cada clúster requiere aproximaciones de intervención adaptadas a sus características particulares, desde programas de apoyo socioeconómico. La comprensión de estos patrones permite trascender aproximaciones genéricas a la retención escolar y desarrollar estrategias focalizadas que aborden las causas específicas de riesgo identificadas en cada perfil poblacional.

Discusión

Este estudio analizó la deserción escolar en Paraguay mediante aprendizaje automático y minería de datos, identificando patrones y factores socioeconómicos clave. Al integrar y procesar grandes volúmenes de datos de fuentes diversas, superó las limitaciones de los análisis tradicionales, proporcionando una visión más completa de la problemática.

La presencia del trabajo infantil emerge como un factor crítico de deserción

en ambos apartados. Mientras que Peralta et al. (Peralta, Misiego Telesca, Benítez Ayala, & Prieto, 2013) identificaron el trabajo juvenil remunerado (35.85%) y los problemas familiares (24.70%) como las principales causas de deserción en Paraguay, vinculadas a la pobreza estructural, el perfil de “alta vulnerabilidad” (clúster 1) en esta investigación muestra una marcada incidencia del 29% de trabajo infantil en los hogares, que se eleva al 46% en zonas rurales. Esto valida la preocupación por el trabajo infantil como un motor crucial del abandono escolar.

Asimismo, las disparidades entre zonas rurales y urbanas son un punto de convergencia fundamental. La Comisión Económica para América Latina y el Caribe (CEPAL) (CEPAL, 2022) ya destacaba que la deserción es significativamente mayor en áreas rurales debido a la escasez de recursos y la necesidad de trabajar. El análisis de agrupamiento realizado refuerza esta premisa al asociar el perfil de “alta vulnerabilidad” y la concentración de factores de riesgo severos directamente con las zonas rurales. La investigación detalla cómo la inasistencia escolar y la escolaridad atrasada son más pronunciadas en el clúster de mayor vulnerabilidad, que a su vez tiene una mayor presencia en áreas rurales.

Finalmente, la metodología de utilizar minería de datos y aprendizaje automático propuesta y empleada en este

estudio, como el algoritmo K-Means, se alinea con las recomendaciones de Abideen (Abideen, Z.U., 2023) sobre el potencial de estas herramientas para descubrir patrones en la deserción escolar e identificar estudiantes en riesgo. La capacidad de generar tres perfiles diferenciados, alta vulnerabilidad, vulnerabilidades moderadas y ventajas socioeconómicas relativas, y de vincularlos directamente con las dimensiones del Índice de Pobreza Multidimensional, va más allá de las estadísticas descriptivas al ofrecer una comprensión matizada y accionable del fenómeno de la deserción.

Conclusiones

Este estudio logró analizar la relación entre la tasa de deserción escolar en Paraguay y el índice de pobreza multidimensional del año 2022 utilizando técnicas de minería de datos y aprendizaje automático, específicamente el algoritmo K-Means para el agrupamiento, se logró identificar patrones y perfiles claros entre los estudiantes de escuelas públicas que tienen entre 10 y 17 años. Esto significa que no solo se confirmó la existencia de una relación del problema, sino que también se organizó a los estudiantes en categorías con riesgos específicos. Por ejemplo, se encontraron grupos con alta vulnerabilidad socioeconómica y otros con desafíos más moderados, lo que permite entender mejor las diferentes situaciones que empujan a los jóvenes a abandonar la escuela.

El estudio identificó patrones de deserción escolar en estudiantes de 10 a 17 años en escuelas públicas de Paraguay, utilizando el algoritmo K-Means para crear dos clústeres: Clúster 1 (alta vulnerabilidad) y Clúster 0 (ventajas socioeconómicas). El clúster 1 mostró dos perfiles: uno de alta vulnerabilidad social y económica, con pobreza, trabajo infantil y alta inasistencia escolar, especialmente en zonas rurales, y otro de vulnerabilidad moderada, con pobreza intermedia y menos trabajo infantil. En cambio, el clúster 0 presentó un perfil de baja vulnerabilidad, con hogares sin pobreza ni trabajo infantil y completa asistencia escolar. Los factores como la ubicación geográfica (rural vs. urbana), modalidad educativa (técnica o científica) y desigualdad de género influyen en la deserción, sugiriendo que, aunque la pobreza extrema es un factor clave, también existen otros determinantes que afectan a estudiantes en contextos más favorables.

El Índice de Pobreza Multidimensional (IPM) del año 2022 se mostró de forma diferenciada entre los perfiles de riesgo identificados. El clúster de alta vulnerabilidad (clúster 1), donde se concentra la mayor inasistencia y escolaridad atrasada, muestra una alta proporción de hogares en situación de pobreza multidimensional, con un 33% de hogares “Pobre No Extremo” y un 14% “Pobre Extremo”. En contrapartida, el clúster con

ventajas socioeconómicas relativas (clúster 0) está compuesto mayormente de 81% de hogares “No Pobre”, aunque también incluye una minoría de hogares en situación de pobreza multidimensional. Esto confirma que el IPM es un factor determinante y que sus distintos niveles están directamente relacionados con la probabilidad de deserción escolar.

La investigación confirma una relación significativa entre los factores socioeconómicos y la probabilidad de deserción escolar en escuelas públicas en Paraguay, nuestra hipótesis apoya la literatura, indicando que estos factores son claves en la deserción escolar (Organización Internacional del Trabajo (OIT), 2017); (Psacharopoulos, 1997); (Instituto Nacional de Estadística (INE), 2022). Mediante el agrupamiento K-Means, se identificaron tres perfiles de riesgo claros. Un grupo de alta vulnerabilidad (clúster 1) se caracteriza por profunda pobreza multidimensional, alta incidencia de trabajo infantil (notablemente en zonas rurales) y significativo atraso escolar e inasistencia. Existen también perfiles con vulnerabilidades moderadas y ventajas socioeconómicas relativas. Estos hallazgos demuestran que, si bien la pobreza es central, la ubicación geográfica (rural/urbana), la modalidad educativa y las disparidades de género también son determinantes en los patrones de deserción (Alkire & Foster, Counting and

Multidimensional Poverty Measurement, 2008); (Alkire & Santos, Acute Multidimensional Poverty: A New Index for Developing Countries, 2010).

Referencias

- World Bank Group. (s.f.). *Poverty Overview*. (World Bank) Recuperado el julio de 2025, de <https://www.worldbank.org/en/topic/poverty/overview>
- Abideen, Z.U. (2023). Analysis of enrollment criteria in secondary schools using machine learning and data mining approach. *Electronics*, 12(3), 694.
- Alkire, S., & Foster, J. (2008). Counting and Multidimensional Poverty Measurement. *Oxford Poverty and Human Development Initiative (OPHI) Working Paper No. 07*. Recuperado el julio de 2025, de <https://ophi.org.uk/publication/WP-07/>
- Alkire, S., & Santos, M. E. (2010). Acute Multidimensional Poverty: A New Index for Developing Countries. *OPHI Working Paper 38*. Obtenido de <https://ophi.org.uk/wp-content/uploads/ophi-wp38.pdf>
- Arthur, D. &. (2007). k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (págs. 1027–1035). New Orleans, LA: Society for Industrial and Applied Mathematics.
- CEPAL. (Noviembre de 2022). *Panorama Social de América Latina 2001 - 2002*. Obtenido de <https://repositorio.cepal.org/server/api/core/bitstreams/596665b2-336e-41df-a0a3-d2c5fc275269/content>
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd ed. O'Reilly Media.
- Gobiernodelparaguay.gov.py. (2024). *Hambre Cero en las Escuelas – Un Paraguay que cuida a los niños*. Recuperado el 2025, de <https://hambrezero.gobiernodelparaguay.gov.py/wp-content/uploads/2025/03/HAMBRE-CERO INFORMACIONES-GENERALES.pdf>
- Heredia Mayorga, H. (2020). Big data en la educación. *SciELO Bolivia*.

- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación* (6a ed. ed.). México: McGraw-Hill.
- Inmon, W. H. (2005). *Building the Data Warehouse* (4ta. Edición ed.). Hoboken.
- Instituto Nacional de Estadística (INE). (Diciembre de 2022). Índice de Pobreza Multidimensional (IPM) 2022. Asunción, Paraguay. Obtenido de https://www.ine.gov.py/publicaciones/sociales/IPM_2022.pdf
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*.
- JetBrains. (2025). PyCharm Community Edition. Obtenido de <https://www.jetbrains.com/pycharm/>
- Kimball , R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd ed. John Wiley & Sons.
- Matplotlib Development Team,. (2023). Matplotlib: Visualization with Python. Obtenido de <https://matplotlib.org/>
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *9th Python in Science Conference*, 51–56.
- NumPy Developers. (2023). NumPy: The fundamental package for scientific computing with Python. Obtenido de <https://numpy.org/>
- Organización Internacional del Trabajo (OIT). (2017). Trabajo infantil: Un problema global. Obtenido de <https://www.ilo.org/es/temas-y-sectores/trabajo-infantil>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Peralta, N., Misiego Telesca, P., Benítez Ayala, M., & Prieto, J. (2013). Deserción en la Educación Media. *Revista Científica Estudios e Investigaciones*.
- Psacharopoulos, G. (1997). Child labor versus educational attainment: Some evidence from Latin America. *Journal of Population Economics*, 10(4), 377–386.

Reyes, G.E., Sosa, D.V., Quispe, M.G., Iraola-Real, I. (2021). Family, economic and pedagogical factors involved in dropping out of school and the consequences for students of a private institution in Lima-Peru. *2021 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, pp. 1–4.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Sabino, C. (2017). *La Historia y su Método: Guía para Estudiantes y Estudiosos de la Historia*. Unión Editorial.

Seaborn Developers. (2025). Seaborn: Statistical data visualization. Obtenido de <https://seaborn.pydata.org/>

Tan, , P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining*. 2nd ed. Pearson.

Tejada Escobar, F., Murrieta Marcillo, R., Villao Santos, F., & Garzón Balcázar, J. (2018). Big Data en la Educación: Beneficios e Impacto de la Analítica de Datos. *SciELO Ecuador*.