

Análisis predictivo de la deserción escolar en áreas rurales de Colombia: identificación de factores de riesgo

Luisa Fernanda Chacón Miranda

Asesor

Javier Ernesto Rodríguez Hernández

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Ingeniería y Tecnología ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Agradecimientos

Agradezco, en primer lugar, a la Universidad Nacional Abierta y a Distancia (UNAD), por ofrecer una educación superior de calidad a tantas personas que, por diversas circunstancias, no pueden asistir de manera presencial a una institución, pero que desean cumplir sus sueños académicos y profesionales.

Expreso también mi gratitud a todos los tutores que me acompañaron a lo largo de esta especialización, por su compromiso con el aprendizaje de los estudiantes y su vocación docente.

Finalmente, extiendo un agradecimiento muy especial al profesor Javier Rodríguez, por su dedicación, orientación y valiosas observaciones durante el desarrollo de este proyecto. Su acompañamiento fue clave para alcanzar este resultado.

Dedicatoria

A mi mamá, por su amor incondicional, su apoyo constante y por ser mi pilar en los momentos más desafiantes. Su ejemplo de fortaleza y dedicación me ha inspirado a seguir adelante cada día.

A mi papá, que partió antes de poder compartir este segundo logro. Fuiste un gran impulsor de la idea de que el desarrollo profesional es la mejor herencia que podemos recibir. Te extraño desde aquí hasta el cielo, y este logro también es tuyo.

A mi compañero de vida, por estar a mi lado en cada paso de este camino, por su paciencia, sus palabras de aliento y por creer en mí incluso cuando yo dudaba.

Y a mi gato, por su silenciosa compañía, sus siestas compartidas y por recordarme que incluso en medio del caos, siempre hay espacio para una pausa, un ronroneo y un momento de calma.

Resumen

La deserción escolar en las zonas rurales de Colombia es un problema estructural que impacta negativamente el desarrollo educativo y social del país. Según datos del Ministerio de Educación Nacional (2023), las tasas de abandono escolar en estas áreas son considerablemente más altas debido a múltiples factores socioeconómicos, familiares y educativos. Según el Departamento Administrativo Nacional de Estadística (DANE), en 2020, la inasistencia escolar en áreas rurales se incrementó al 30,1%, en comparación con el 4,8% registrado en 2019, evidenciando una marcada disparidad educativa en estas regiones.

Este proyecto de investigación tiene como objetivo realizar un análisis cuantitativo integral para identificar patrones de deserción y desarrollar modelos predictivos que permitan anticipar casos de abandono escolar. Para ello, se utilizarán diversas fuentes de datos oficiales del Ministerio de Educación Nacional, disponibles en su portal de datos abiertos (<https://www.datos.gov.co/>), incluyendo bases de datos sobre matrícula, infraestructura educativa, indicadores de primera infancia y programas de alimentación escolar.

El estudio se desarrolló en un período de 12 meses bajo un enfoque cuantitativo, combinando análisis descriptivo y predictivo. En la primera fase (meses 1-6), se llevó a cabo una revisión bibliográfica exhaustiva basada en autores como Albor-Chadid et al. (2019), Aravena Domich & Pérez Campillo (2022) y Cerón et al. (2024), quienes han abordado distintos aspectos de la deserción escolar en Colombia. En la fase de análisis (meses 7-10), se implementaron técnicas de machine learning, como árboles de decisión y regresión logística, para identificar patrones de deserción. Se trabajó con variables socioeconómicas, demográficas y educativas, extraídas de bases de datos clave, tales como:

- Infraestructura educativa.

- Datos de establecimientos educativos.
- Estadísticas por departamentos y municipios.
- Programas de alimentación escolar
- Cifras de casos de violencia por conflicto armado.
- Índice de pobreza multidimensional

Y en la tercera fase o fase final (mes 11-12) se completó la documentación de los hallazgos y se presentó el producto final del proyecto. Los resultados obtenidos incluyeron:

- Se identificaron perfiles de estudiantes en riesgo de deserción, caracterizados por bajos niveles de aprobación y alta repetencia, influenciados por factores socioeconómicos y educativos.
- Se elaboraron mapas de riesgo que evidencian mayores tasas de deserción en departamentos como Guainía, Caquetá y Vichada, ofreciendo una herramienta visual para la toma de decisiones.
- Se propusieron recomendaciones de política pública orientadas a mejorar la calidad educativa y mitigar la pobreza multidimensional como estrategias clave.

Este trabajo representa una contribución significativa al campo de la educación rural, al ofrecer evidencia concreta para intervenciones específicas. Los hallazgos serán de utilidad para diversas entidades, incluyendo el Ministerio de Educación Nacional, Secretarías de Educación Departamentales, organizaciones no gubernamentales enfocadas en educación rural e investigadores académicos interesados en políticas educativas.

El proyecto se enmarca en la línea de investigación de Big Data y Machine Learning de la Escuela de Ciencias Básicas, Ingeniería y Tecnología (ECBTI).

Palabras claves: Deserción, escolaridad, rural, análisis de datos, equidad.

Abstract

School dropout in rural areas of Colombia is a structural problem that negatively impacts the country's educational and social development. According to data from the Ministry of National Education (2023), dropout rates in these areas are considerably higher due to multiple socioeconomic, family, and educational factors. According to the National Administrative Department of Statistics (DANE), in 2020, school absenteeism in rural areas rose to 30.1%, compared to 4.8% in 2019, highlighting a marked educational disparity in these regions.

This research project aims to carry out a comprehensive quantitative analysis to identify dropout patterns and develop predictive models that can anticipate cases of school abandonment. To this end, various official data sources from the Ministry of National Education, available on its open data portal (<https://www.datos.gov.co/>), will be used. These include databases on enrollment, educational infrastructure, early childhood indicators, and school feeding programs.

The study was conducted over a 12-month period under a quantitative approach, combining descriptive and predictive analysis. In the first phase (months 1–6), an extensive literature review was carried out, based on authors such as Albor-Chadid et al. (2019), Aravena Domich & Pérez Campillo (2022), and Cerón et al. (2024), who have addressed different aspects of school dropout in Colombia. In the analysis phase (months 7–10), machine learning techniques such as decision trees and logistic regression were implemented to identify dropout patterns. The analysis included socioeconomic, demographic, and educational variables extracted from key databases, such as:

- Educational infrastructure.
- Data on educational institutions.
- Statistics by departments and municipalities.

- School feeding programs.
- Figures on violence due to armed conflict.
- Multidimensional poverty index.

In the third and final phase (months 11–12), the documentation of findings was completed, and the final project product was presented. The results included:

- Identification of student profiles at risk of dropping out, characterized by low academic achievement and high repetition rates, influenced by socioeconomic and educational factors.
- Development of risk maps showing higher dropout rates in departments such as Guainía, Caquetá, and Vichada, providing a visual tool for decision-making.
- Policy recommendations aimed at improving educational quality and mitigating multidimensional poverty as key strategies were proposed.

This work represents a significant contribution to the field of rural education, offering concrete evidence for targeted interventions. The findings will be useful for various entities, including the Ministry of National Education, Departmental Education Secretariats, non-governmental organizations focused on rural education, and academic researchers interested in education policy.

The project is part of the Big Data and Machine Learning research line of the School of Basic Sciences, Engineering, and Technology (ECBTI).

Keywords: Dropout, schooling, rural, data analysis, equity.

Contenido

Introducción	14
Descripción del Problema	16
Planteamiento del Problema	16
Sistematización del Problema	18
Dimensiones Principales.....	19
Componente Analítico.	19
Aspecto predictivo.	19
Impacto del Conflicto Armado.	20
Implicaciones Políticas.	20
Disparidades Regionales.....	20
Justificación	21
Objetivos.....	24
Objetivo General.....	24
Objetivos Específicos.....	24
Marcos de Referencia	25
Estado del Arte.....	25
Antecedentes Internacionales	26
Antecedentes Nacionales	27
Antecedentes Regionales	28
Brechas y Oportunidades Identificadas	30
Marco Contextual.....	31
Situación Actual de la Educación Rural	31

Políticas y Programas Educativos.....	31
Factores Socioeconómicos.....	32
Vinculación con la Deserción Escolar	33
Marco Teórico.....	33
Teoría del Capital Social (Putnam, 2000).....	33
Teoría del Conflicto Social (Marx & Engels, 1848).....	34
Teoría del Comportamiento Planificado (Ajzen, 1991).....	34
Marco Conceptual.....	35
Marco Normativo.....	35
Normativa Nacional.....	36
Políticas Públicas Relevantes	37
Normativa Internacional	37
Normativa sobre Datos Abiertos y Transparencia.....	38
Metodología	39
Método	39
Tipo de Estudio.....	39
Recolección de Datos.....	39
Fuentes Secundarias.....	39
Procesamiento de Datos.....	40
Instrumentos Principales.....	41
Matriz de Recolección de Datos.	42
Fases del Proyecto.....	44
Cronograma de Actividades.....	45

Resultados	47
Contexto Regional: Indicadores Educativos, Violencia y Pobreza Multidimensional	47
Departamentos y Municipios con Mayores Casos de Violencia	56
Índice de Pobreza Multidimensional (IPM)	59
Estadísticas Descriptivas por Variable para el Año 2024.....	60
Brechas Rurales contra Cabeceras Municipales	61
Implicaciones para Políticas Públicas y Conclusiones	62
Procedimientos de Análisis.....	62
Integración y Limpieza de Datos	62
Análisis Descriptivo.....	63
Análisis de Correlación.....	63
Modelos de Regresión Lineal	63
Árboles de Decisión.....	64
Visualización de Resultados	64
Hallazgos Clave	64
Interpretación y Discusión	76
Conclusiones	81
Recomendaciones	86
Reflexiones Finales.....	91
Referencias Bibliográficas	93
Apéndices.....	101

Lista de Tablas

Tabla 1 <i>Matriz de Recolección de Datos por Variable</i>	42
Tabla 2 <i>Cronograma de Actividades</i>	45
Tabla 3 <i>Resumen de Correlaciones entre Variables de Conflicto y Académicas</i>	67
Tabla 4 <i>Variables Educativas y su Relación con Condiciones Sociales</i>	69
Tabla 5 <i>Modelo de Regresión Lineal – Desaparición Forzada</i>	70
Tabla 6 <i>Modelo de Regresión Lineal – Reclutamiento Forzado</i>	71
Tabla 7 <i>Modelo de Regresión Lineal – Violencia Sexual</i>	72
Tabla 8 <i>Contraste entre Objetivos y Resultados</i>	84

Lista de Figuras

Figura 1 <i>Top 10 Departamentos con Mayor Tasa de Deserción</i>	48
Figura 2 <i>Top 10 Municipios con Mayor Tasa de Deserción</i>	49
Figura 3 <i>Top 10 Departamentos con Menor Tasa de Deserción</i>	49
Figura 4 <i>Top 10 Departamentos con Mayor Tasa de Repitencia</i>	50
Figura 5 <i>Top 10 Municipios con Mayor Tasa de Repitencia</i>	51
Figura 6 <i>Top 10 Departamentos con Menor Tasa de Repitencia</i>	51
Figura 7 <i>Top 10 Departamentos con Menor Tasa de Matrícula</i>	52
Figura 8 <i>Top 10 Municipios con Menores Tasas de Matrícula</i>	53
Figura 9 <i>Top 10 Departamentos con Mayores Tasas de Matrícula</i>	53
Figura 10 <i>Top 10 Departamentos con Mayor Tasa de Aprobación</i>	54
Figura 11 <i>Top 10 Municipios con Mayor Tasa de Aprobación</i>	55
Figura 12 <i>Top 10 Departamentos con Menor Tasa de Aprobación</i>	55
Figura 13 <i>Top 10 Departamentos con más Casos de Violencia Sexual</i>	56
Figura 14 <i>Top 10 Departamentos con más Casos de Reclutamiento Forzado</i>	57
Figura 15 <i>Top 10 Municipios con Mayores Casos de Reclutamiento</i>	57
Figura 16 <i>Top 10 Departamentos con más Casos de Desaparición Forzada</i>	58
Figura 17 <i>Top 10 Municipios con más Casos de Desaparición Forzada</i>	59
Figura 18 <i>Mapa de Calor de Correlaciones entre Variables</i>	68

Lista de Apéndices

Apéndice A *Visual del Código Python con Análisis de Datos* 101

Apéndice B *Tablero Power BI para Análisis de Datos* 101

Introducción

La deserción escolar en las zonas rurales de Colombia sigue siendo un desafío significativo para el desarrollo educativo y social del país. Según datos recientes del Ministerio de Educación Nacional, en 2023, un total de 473.780 menores abandonaron las aulas, superando la cifra registrada en 2022, que fue de 337.104. La tasa de deserción es mayor en las zonas rurales, alcanzando el 5,4%, en comparación con el 3,6% en las áreas urbanas, lo que evidencia una brecha educativa persistente entre ambas áreas (Sectorial, 2024).

Además, en 2022, el 23,7% de los jóvenes en zonas rurales no asistió a ninguna institución educativa, reflejando una disminución en comparación con el 26,9% registrado en 2021. Sin embargo, esta cifra sigue siendo alarmante y pone de manifiesto las dificultades de acceso a la educación en estas áreas (La República, 2023). Estos datos subrayan la necesidad de implementar políticas públicas más efectivas y focalizadas que aborden las causas subyacentes de la deserción escolar en las zonas rurales de Colombia.

Este fenómeno no solo compromete el futuro de miles de niños y jóvenes colombianos, sino que también perpetúa ciclos de pobreza y exclusión social. Como destacan Gómez Triana (2016) y Utria (2023), factores como la precariedad socioeconómica, el impacto del conflicto armado y las limitaciones estructurales del sistema educativo rural han generado un entorno donde el acceso y la permanencia en la escuela se ven gravemente afectados.

En respuesta a esta problemática, el presente proyecto busca aportar una visión integral mediante el uso de herramientas analíticas avanzadas. La disponibilidad de bases de datos oficiales del Ministerio de Educación ofrece una oportunidad única para realizar un análisis cuantitativo profundo, permitiendo identificar patrones de deserción y sus principales factores predictivos.

Esta investigación no solo se justifica por su relevancia académica, sino también por su impacto potencial en la formulación de políticas públicas más efectivas. Al combinar técnicas de análisis descriptivo y predictivo, se pretende generar conocimiento accionable que facilite la implementación de estrategias más precisas y focalizadas por parte del Ministerio de Educación, las administraciones locales y organizaciones dedicadas a la educación en zonas rurales.

Descripción del Problema

Planteamiento del Problema

La deserción escolar en las zonas rurales de Colombia representa un desafío estructural que afecta tanto el desarrollo educativo como las dinámicas sociales y económicas del país. Según datos recientes del Ministerio de Educación Nacional (MEN, 2023), la tasa de deserción en educación secundaria en áreas rurales alcanza el 5.6%, superando en 1.4 puntos porcentuales el 4.2% registrado en zonas urbanas. Esta brecha equivale a aproximadamente 65,000 estudiantes que abandonan el sistema educativo cada año en contextos rurales, lo que no solo restringe sus oportunidades de desarrollo personal, sino que también perpetúa los ciclos intergeneracionales de pobreza y desigualdad.

Las causas de esta problemática son variadas y están interconectadas, reflejando la complejidad de los entornos rurales. El Sistema de Información para el Monitoreo, la Prevención y el Análisis de la Deserción Escolar (SIMPADE, 2023) indica que el 42% de los casos de deserción en estas zonas se atribuye a factores económicos, como la necesidad de los estudiantes de trabajar para apoyar a sus familias; el 27% está relacionado con barreras de accesibilidad geográfica, incluyendo largas distancias y falta de transporte; el 18% se vincula a dinámicas familiares, como el bajo nivel educativo de los padres; y el 13% restante responde a deficiencias institucionales, como la escasez de docentes o materiales pedagógicos adecuados. Este panorama evidencia que la deserción escolar trasciende el ámbito educativo y se entrelaza con condiciones estructurales de inequidad territorial y social.

Investigadores como Gómez Triana (2020) y Utria (2023) señalan que la deserción en zonas rurales se ve agravada por factores históricos y contextuales, como el impacto persistente del conflicto armado y las limitaciones del sistema educativo en estas regiones. Por su parte, el

Departamento Nacional de Planeación (DNP, 2023) ha identificado que en municipios con un índice de pobreza multidimensional superior al 60%, la probabilidad de abandono escolar es 2.3 veces mayor que en aquellos con mejores condiciones socioeconómicas. Estas disparidades subrayan la urgencia de abordar el problema desde una perspectiva integral que combine análisis cuantitativo y estrategias focalizadas.

Frente a esta situación, el presente proyecto se propone responder a la necesidad de comprender en profundidad las dinámicas de la deserción escolar mediante el uso de herramientas analíticas avanzadas. La abundancia de datos oficiales provenientes del Ministerio de Educación Nacional y otras fuentes gubernamentales ofrece una oportunidad valiosa para realizar un análisis exhaustivo que identifique patrones, correlaciones y factores predictivos asociados al abandono escolar. Este enfoque no solo busca diagnosticar el problema, sino también proponer soluciones basadas en evidencia que puedan orientar intervenciones efectivas.

Por lo anterior, el estudio busca responder a la siguiente pregunta de investigación:

- ¿Cuáles son los principales factores socioeconómicos, educativos y contextuales que predicen la deserción escolar en zonas rurales de Colombia, y cómo pueden integrarse en un modelo predictivo para anticipar casos de riesgo y orientar intervenciones educativas efectivas?

Para abordar esta pregunta, el proyecto adoptará un enfoque cuantitativo que considere las siguientes dimensiones clave:

- Factores socioeconómicos y familiares: Nivel de ingresos, condiciones de vivienda, y educación de los padres como determinantes del abandono escolar.
- Impacto del conflicto armado: Efectos del desplazamiento forzado, la violencia y el reclutamiento infantil en la permanencia educativa.

- Infraestructura educativa: Influencia de la escasez de docentes, el estado de las escuelas y la disponibilidad de recursos pedagógicos.
- Barreras geográficas: Dificultades de acceso a las instituciones debido a la dispersión territorial y la falta de vías adecuadas.
- Factores culturales y comunitarios: Percepciones sociales sobre la educación y la presión para que los menores trabajen en actividades económicas.

Este análisis multidimensional permitirá generar conocimientos sólidos que sirvan de base para diseñar estrategias de intervención y políticas públicas orientadas a reducir la deserción escolar en las zonas rurales, promoviendo así una educación más equitativa y accesible para esta población.

Sistematización del Problema

La sistematización del problema de la deserción escolar en zonas rurales de Colombia resulta esencial para desentrañar su complejidad y diseñar soluciones efectivas que respondan a las necesidades reales de esta población. Este proceso implica descomponer el fenómeno en sus elementos constitutivos, identificar las interrelaciones entre ellos y priorizar las áreas de análisis que permitan generar un impacto significativo. Al hacerlo, se busca trascender las descripciones superficiales y avanzar hacia una comprensión profunda que combine datos cuantitativos, evidencia empírica y enfoques teóricos. Esta metodología estructurada no solo facilita la identificación de los factores más críticos, sino que también sienta las bases para el desarrollo de modelos predictivos y estrategias de intervención basadas en hallazgos sólidos.

En este sentido, la sistematización del problema se convierte en un paso previo indispensable para orientar el estudio y garantizar que los resultados sean aplicables en el ámbito práctico. Al estructurar la problemática en dimensiones específicas, como las analíticas,

predictivas, sociopolíticas y regionales, se asegura que el análisis aborde tanto las causas inmediatas como los factores estructurales subyacentes. Además, este enfoque permite formular preguntas orientadoras que guíen la investigación de manera coherente y sistemática, promoviendo un diálogo entre los datos disponibles y las necesidades de las comunidades rurales. Así, la sistematización no solo organiza el conocimiento, sino que también lo transforma en una herramienta accionable para la toma de decisiones.

A partir de esta base, se plantean las siguientes preguntas orientadoras para estructurar el análisis de la deserción escolar en zonas rurales:

- ¿Cuáles son los factores socioeconómicos, familiares y educativos más determinantes en la deserción escolar rural?
- ¿Cómo se interrelacionan estos factores para aumentar el riesgo de abandono escolar?

Dimensiones Principales

Componente Analítico.

- ¿Qué patrones y tendencias emergen del análisis de los datos oficiales del Ministerio de Educación y otras fuentes gubernamentales?
- ¿Cómo varían estos patrones según las características demográficas, geográficas e institucionales de las zonas rurales?

Aspecto predictivo.

- ¿Qué métodos de análisis predictivo ofrecen mayor precisión para identificar estudiantes en riesgo de deserción?
- ¿De qué manera pueden estos modelos optimizar las estrategias de intervención temprana en contextos rurales?

Impacto del Conflicto Armado.

- ¿Cómo ha influido el conflicto armado en las tasas de deserción escolar en diferentes regiones rurales?
- ¿Qué efectos específicos tienen el desplazamiento y la violencia en la permanencia educativa?

Implicaciones Políticas.

- ¿Qué políticas públicas podrían derivarse de los resultados del análisis cuantitativo para reducir la deserción?
- ¿Cómo pueden estas políticas adaptarse a las particularidades de los contextos rurales para maximizar su impacto?

Disparidades Regionales.

- ¿Qué diferencias existen en las tasas y causas de deserción entre las distintas regiones rurales de Colombia?
- ¿Cómo pueden estas disparidades guiar la distribución de recursos y la priorización de intervenciones educativas?

La exploración de estas preguntas y dimensiones permitirá construir un marco analítico robusto que no solo identifique las causas de la deserción escolar, sino que también ofrezca soluciones prácticas y contextualizadas para mitigar este problema en las zonas rurales.

Justificación

Este proyecto de investigación sobre la deserción escolar en zonas rurales de Colombia se justifica plenamente desde diversas perspectivas que confluyen en su relevancia social, política, académica y técnica. En primer lugar, se trata de una propuesta factible, sustentada en la amplia disponibilidad de datos abiertos y oficiales proporcionados por el Ministerio de Educación Nacional, el DANE y otras entidades gubernamentales. Donde, se cuenta con acceso a estas bases de datos, así como con las competencias técnicas y herramientas tecnológicas necesarias para su procesamiento y análisis, empleando metodologías propias de la ciencia de datos. Además, el tiempo estimado para la ejecución del estudio, doce meses, es adecuado para desarrollar cada una de las fases planificadas, asegurando su culminación exitosa.

La pertinencia de esta investigación radica en su enfoque sobre una problemática educativa crítica y actual. Según datos del Ministerio de Educación Nacional (2023), la tasa de deserción escolar en zonas rurales supera en 1.4 puntos porcentuales a la de las zonas urbanas, afectando a cerca de 65,000 estudiantes al año. Este fenómeno no solo representa una pérdida de oportunidades individuales, sino que obstaculiza el cumplimiento de metas globales como el Objetivo de Desarrollo Sostenible, centrado en garantizar una educación inclusiva y equitativa. Asimismo, se alinea con las prioridades establecidas en el Plan Nacional de Desarrollo 2022–2026, que busca reducir las brechas educativas territoriales. Por tanto, el estudio responde a una necesidad urgente y estratégica en el contexto colombiano actual.

El momento en que se plantea esta investigación resulta especialmente propicio dada la preocupación gubernamental por la equidad educativa lo que ha conllevado a la consolidación de plataformas como el Sistema de Información para el Monitoreo, la Prevención y el Análisis de la Deserción Escolar (SIMPADE), que recopilan información antes dispersa. Esta coyuntura

constituye una oportunidad única para realizar análisis cuantitativos más profundos y con mayor precisión. Como lo expresa Carvajal Valderrama (2023), existe una verdadera ventana de oportunidad para transformar datos en insumos relevantes para la formulación de políticas públicas más eficaces.

Los aportes de esta investigación trascienden el ámbito académico y se proyectan hacia aplicaciones prácticas de gran impacto. A nivel institucional, los resultados podrán ser utilizados por escuelas rurales para identificar tempranamente a estudiantes en riesgo de abandonar el sistema escolar. A nivel territorial, permitirán a las secretarías de educación focalizar recursos y diseñar estrategias diferenciales con base en las particularidades de cada región. A nivel nacional, constituirán un insumo valioso para que el Ministerio de Educación fundamente políticas públicas orientadas a garantizar la permanencia escolar en contextos rurales.

Desde el plano teórico, el estudio enriquecerá el conocimiento existente sobre la deserción escolar en Colombia mediante la incorporación de técnicas analíticas avanzadas que han sido escasamente utilizadas en investigaciones previas. Tal como indican Velasco Quintero y Castro Garzón (2023), gran parte de los estudios existentes han privilegiado metodologías cualitativas o análisis estadísticos básicos, dejando un vacío importante en la aplicación de modelos predictivos y herramientas de machine learning. Esta investigación se propone llenar ese vacío, explorando la interacción entre factores educativos, socioeconómicos y territoriales a través de enfoques innovadores.

En cuanto a su utilidad metodológica, el enfoque adoptado representa una contribución significativa al campo del análisis educativo en Colombia. La implementación de algoritmos como árboles de decisión y regresión logística sobre bases de datos oficiales marcará un hito replicable en futuros estudios relacionados con fenómenos educativos. Los procedimientos de

limpieza, integración y análisis de datos desarrollados durante el proyecto constituirán un referente metodológico útil para otros investigadores interesados en la aplicación de la ciencia de datos al ámbito educativo.

En suma, esta investigación no solo es viable y oportuna, sino que tiene el potencial de generar conocimientos relevantes y aplicables que contribuirán directamente a mejorar la permanencia escolar en zonas rurales. Con ello, se espera aportar a la reducción de las desigualdades territoriales en educación y al fortalecimiento del derecho fundamental a la educación para niños, niñas y adolescentes que viven en las regiones más vulnerables del país.

Objetivos

Objetivo General

Analizar los factores predictivos de la deserción escolar en zonas rurales de Colombia mediante técnicas de minería de datos y aprendizaje automático, para diseñar un modelo que identifique patrones de riesgo y proponga estrategias de intervención focalizadas según perfiles de vulnerabilidad.

Objetivos Específicos

Examinar los factores socioeconómicos, familiares y educativos que influyen en la deserción escolar en zonas rurales, categorizándolos según su nivel de impacto mediante análisis estadístico multivariado.

Comparar patrones diferenciales de deserción escolar entre regiones geográficas y grupos poblacionales, identificando correlaciones significativas con variables contextuales como infraestructura educativa, conectividad y presencia institucional.

Construir modelos predictivos mediante técnicas de aprendizaje automático que permitan anticipar situaciones de riesgo de abandono escolar con base en variables cuantificables extraídas de fuentes oficiales.

Formular recomendaciones estratégicas basadas en evidencia cuantitativa para el diseño de políticas públicas e intervenciones diferenciadas según perfiles de riesgo identificados en los modelos predictivos.

Marcos de Referencia

Estado del Arte

La deserción escolar en zonas rurales representa uno de los desafíos más persistentes y complejos en el ámbito educativo, tanto a nivel global como en contextos específicos como el colombiano. Este fenómeno, influenciado por una multiplicidad de factores socioeconómicos, geográficos, culturales y políticos, no solo limita el acceso a la educación como un derecho fundamental, sino que perpetúa ciclos de pobreza, desigualdad y exclusión social. En los últimos años, la investigación sobre este tema ha ganado relevancia debido a la necesidad de comprender sus causas estructurales y de diseñar estrategias efectivas para contrarrestarlo. La incorporación de nuevas perspectivas y estudios recientes permite enriquecer el análisis, aportando una visión más integral que conecta experiencias internacionales con realidades locales. Este estado del arte busca revisar de manera sistemática los antecedentes más significativos en diferentes escalas — internacional, nacional y regional—, así como identificar las brechas y oportunidades que orienten futuras intervenciones. Al hacerlo, se destacan investigaciones recientes que abordan desde los efectos de crisis ambientales y conflictos armados hasta el uso de enfoques innovadores como el aprendizaje automático, los cuales ofrecen nuevas herramientas para enfrentar este problema.

La importancia de este ejercicio radica en su capacidad para visibilizar cómo la deserción escolar en entornos rurales trasciende las barreras meramente educativas y se entrelaza con dinámicas sociales más amplias. En un mundo cada vez más interconectado, las lecciones aprendidas en contextos diversos pueden iluminar soluciones aplicables a Colombia, mientras que los datos nacionales y regionales permiten aterrizar estas ideas en la realidad específica del país. A continuación, se presenta una síntesis estructurada de los hallazgos más relevantes,

integrando tanto referencias clásicas como aportes recientes que fortalecen la comprensión del fenómeno.

Antecedentes Internacionales

A nivel internacional, la deserción escolar en contextos rurales sigue siendo un desafío significativo, exacerbado por factores socioeconómicos, culturales y geográficos. Estudios recientes han arrojado luz sobre las complejidades de este fenómeno en diferentes regiones del mundo, ofreciendo perspectivas útiles para el caso colombiano. Por ejemplo, según un informe de UNICEF, más de 420.000 niños en Brasil, Colombia y Perú se vieron gravemente afectados por una sequía sin precedentes en la Amazonia, lo que resultó en malnutrición, abandono escolar y aislamiento de comunidades. En Brasil, más de 1.700 escuelas y 760 centros de salud cerraron, afectando a medio millón de niños. Esta crisis evidenció cómo factores ambientales pueden incidir directamente en la continuidad educativa en zonas rurales (UNICEF, 2024).

En Irán, Eslamian et al. (2023) llevaron a cabo un estudio cualitativo para explorar las razones del abandono escolar en áreas rurales. Sus hallazgos revelaron que la pobreza, la falta de acceso a servicios educativos de calidad y las expectativas familiares son determinantes clave en la decisión de los estudiantes de dejar la escuela. Este análisis subraya la necesidad de entender las barreras específicas que enfrentan las comunidades rurales en distintos contextos culturales, un enfoque que puede adaptarse para identificar factores similares en Colombia.

Por su parte, el informe de Bradley, M. (2024), examinó las elevadas tasas de deserción escolar en China rural, uno de los países con mayor población rural del mundo. El estudio destaca que, pese a las políticas gubernamentales para mejorar la educación rural, persisten desafíos como la migración laboral de los padres, la insuficiencia de infraestructura educativa y la presión económica sobre las familias. Estos factores contribuyen a tasas de abandono que

limitan el desarrollo educativo, ofreciendo lecciones sobre la importancia de políticas integrales que aborden las realidades socioeconómicas de las zonas rurales, aplicables al diseño de intervenciones en Colombia.

Antecedentes Nacionales

En Colombia, la deserción escolar en áreas rurales ha sido objeto de creciente atención en los últimos años, especialmente tras eventos disruptivos como la pandemia de COVID-19. Estudios recientes han profundizado en los factores que impulsan este fenómeno y en las respuestas institucionales para mitigarlo.

Según el Laboratorio de Economía de la Educación (LEE) de la Universidad Javeriana, en 2022, el 23,7% de los jóvenes en zonas rurales no asistió a instituciones educativas, evidenciando una disminución respecto al 26,9% registrado en 2021, pero aun reflejando una problemática significativa (LEE, 2023). Además, menos de la mitad de los estudiantes que ingresan a primero de primaria en zonas rurales logran culminar sus estudios de bachillerato, lo que refleja una preocupante tasa de deserción en estos contextos (LEE, 2024).

Morales Vega (2021), en su investigación realizada en los municipios de Chinú y Ortega y publicada por la Universidad Javeriana, analizó las causas de la deserción escolar en comunidades rurales durante la pandemia. Los resultados identificaron la falta de acceso a servicios educativos de calidad, la pobreza extrema y la necesidad de los estudiantes de trabajar para apoyar a sus familias como principales impulsores del abandono. Este estudio establece una base sólida para comprender las realidades locales que afectan la permanencia escolar en el país.

Complementando estos hallazgos, United Way Colombia (2022) publicó un informe que evalúa el impacto de la pandemia en la deserción escolar rural. El análisis señala que la crisis sanitaria agravó las desigualdades educativas preexistentes, incrementando las tasas de abandono

debido a la falta de acceso a herramientas de educación virtual y al aumento de la presión económica sobre las familias. Este trabajo destaca la urgencia de estrategias que respondan a los desafíos coyunturales y estructurales que enfrentan las zonas rurales colombianas.

Asimismo, el Ministerio de Educación Nacional (2022) identifica múltiples factores que inciden en la deserción escolar en Colombia, entre ellos, la pobreza, la falta de infraestructura adecuada, la escasez de docentes y la ausencia de programas de acompañamiento y apoyo psicosocial. Estos factores afectan de manera diferencial a los estudiantes de zonas rurales, quienes deben enfrentar barreras adicionales como largas distancias para acceder a la escuela y condiciones precarias en las instituciones educativas.

De igual manera, una revisión realizada dentro de la Universidad Militar Nueva Granada analiza cómo la violencia y el desplazamiento forzado afectan la cobertura educativa, la calidad de la enseñanza y las tasas de deserción escolar (Palacios Tovar, M., 2017). La investigación destaca que la intensificación del conflicto armado ha exacerbado las desigualdades educativas en las zonas rurales, limitando las oportunidades de acceso y permanencia en el sistema educativo.

Estos estudios ofrecen un contexto esencial para esta investigación, al evidenciar cómo factores estructurales y eventos recientes interactúan para influir en la deserción escolar, subrayando la necesidad de enfoques multidimensionales en el análisis y la intervención.

Antecedentes Regionales

Las disparidades regionales en las tasas de deserción escolar en Colombia son marcadas, reflejando las diferencias socioeconómicas y geográficas entre las zonas del país. Investigaciones recientes han mapeado estas variaciones y explorado sus causas subyacentes, proporcionando un marco para intervenciones focalizadas.

El Departamento Administrativo Nacional de Estadística (DANE, 2023) publicó estadísticas detalladas sobre la deserción escolar por departamento, revelando tasas significativamente más altas en regiones rurales y periféricas. Por ejemplo, departamentos como Guainía (6.7%) y Vichada (6.1%) superan ampliamente el promedio nacional de 3.49%, lo que indica una concentración de riesgos en áreas con mayores privaciones. Estos datos enfatizan la necesidad de políticas educativas que prioricen las regiones más afectadas.

Además, SITEAL (2022) ofrece un análisis regional detallado de las tendencias y patrones de deserción en Colombia, destacando factores contextuales como la dispersión geográfica, la falta de infraestructura y la influencia del conflicto armado. Aunque este estudio abarca un periodo amplio, sus hallazgos recientes son valiosos para identificar las disparidades regionales y los determinantes específicos que contribuyen al abandono escolar en cada zona, proporcionando una base para estrategias adaptadas a las necesidades locales.

También, un estudio publicado en la revista *Perfiles Educativos* (2021) señala que las instituciones educativas rurales presentan condiciones de alta vulnerabilidad debido a su ubicación geográfica, el abandono estatal y la presencia de actores armados. Estas condiciones generan un entorno de inestabilidad que desmotiva la continuidad educativa y propicia la deserción escolar.

Por otro lado, la investigación de Sosa Chinome y Vargas Rodríguez (2022) analiza las consecuencias de la deserción escolar en niños y jóvenes víctimas del conflicto armado, destacando sus implicaciones sociales y económicas a mediano y largo plazo. El estudio enfatiza que la falta de acceso a la educación limita las oportunidades de empleo y desarrollo personal, aumentando el riesgo de que los jóvenes sean reclutados por grupos armados ilegales o se involucren en economías informales de subsistencia.

Estos antecedentes regionales son fundamentales para comprender las variaciones de la deserción dentro de Colombia, destacando la importancia de considerar el contexto geográfico y social en el diseño de soluciones efectivas.

Brechas y Oportunidades Identificadas

La revisión de estos estudios evidencia que la deserción escolar en las zonas rurales de Colombia es un problema multifactorial que puede llegar a estar influenciado por factores socioeconómicos, geográficos y estructurales. La interacción de estos factores contribuye a perpetuar la desigualdad y la exclusión en comunidades rurales, limitando el acceso a oportunidades educativas y de desarrollo. Estos hallazgos resaltan la urgencia de diseñar e implementar políticas públicas integrales que aborden estas problemáticas de manera coordinada, con el fin de garantizar el derecho a una educación de calidad para todos los niños y jóvenes en contextos rurales.

Por último, del análisis de la literatura existente se identifican las siguientes brechas investigativas que justifican nuestro proyecto:

- Escasez de estudios que apliquen técnicas avanzadas de análisis predictivo específicamente al contexto rural colombiano.
- Limitada integración de múltiples fuentes de datos oficiales para obtener una visión integral del fenómeno.
- Insuficiente desarrollo de modelos predictivos que consideren variables territoriales específicas como la intensidad del conflicto armado, la presencia de economías ilícitas o la dotación de infraestructura.
- Ausencia de herramientas analíticas que permitan segmentar perfiles de riesgo según características regionales y poblacionales para la focalización de intervenciones.

La presente investigación busca contribuir a llenar estas brechas mediante la aplicación de técnicas avanzadas de análisis cuantitativo que integren múltiples fuentes de datos oficiales para generar modelos predictivos específicos para el contexto rural colombiano.

Marco Contextual

La educación rural en Colombia se refiere al sistema educativo que opera en áreas de baja densidad poblacional, caracterizadas por su dispersión geográfica y una economía predominantemente agrícola (DANE, 2022). Este sector educativo atiende a aproximadamente el 20% de la población estudiantil del país, pero enfrenta desafíos estructurales que limitan su calidad y continuidad, siendo la deserción escolar uno de los problemas más críticos.

Situación Actual de la Educación Rural

El análisis de la deserción escolar en Colombia debe entenderse dentro del contexto socioeconómico y político del país. Según el Ministerio de Educación Nacional (2022), las tasas de abandono escolar en las zonas rurales son considerablemente más altas que en los entornos urbanos, debido a múltiples factores interrelacionados que dificultan la permanencia de los estudiantes en el sistema educativo.

Este fenómeno se ve agravado por la precariedad de la infraestructura: solo el 30% de las escuelas rurales cuentan con acceso a internet (DANE, 2023), y muchas carecen de recursos básicos como bibliotecas o laboratorios. Además, la calidad educativa es baja, reflejada en tasas de repitencia que alcanzan el 19.4% en regiones como Guainía (MEN, 2023).

Políticas y Programas Educativos

El gobierno colombiano ha implementado iniciativas como el "Pacto por la Educación Rural" (2019), que busca mejorar la infraestructura y la formación docente en estas zonas. Sin embargo, su efectividad ha sido limitada por la falta de recursos y la dificultad de llegar a

comunidades dispersas. Programas como "Todos a Aprender" han mejorado algunos indicadores de calidad, pero no han logrado reducir significativamente la deserción, que sigue siendo un desafío persistente.

Factores Socioeconómicos

La pobreza multidimensional es un determinante clave en la educación rural. En departamentos como Vichada, donde el Índice de Pobreza Multidimensional (IPM) llega al 70.2%, la deserción escolar alcanza el 6.1% (DANE, 2023). La necesidad de que los estudiantes trabajen en actividades agrícolas o familiares, junto con la baja valoración de la educación en algunas comunidades, contribuye al abandono escolar.

Asimismo, uno de los principales factores que ha impactado la educación rural en Colombia es el conflicto armado. Como señala Utria (2023), las comunidades afectadas por la violencia presentan mayores tasas de deserción debido a desplazamientos forzados, inseguridad y la necesidad de priorizar la subsistencia sobre la educación formal. En muchas regiones, la presencia de grupos armados limita la asistencia escolar y afecta la estabilidad del sistema educativo.

Desde una perspectiva económica, Carvajal Valderrama (2023) documentó cómo prácticas como las aspersiones aéreas con glifosato han exacerbado las dificultades educativas en zonas rurales, afectando la salud de las comunidades y su acceso a medios de subsistencia. Además, la pobreza estructural, el desempleo y la falta de infraestructura básica continúan limitando el acceso a servicios educativos de calidad, lo que aumenta el riesgo de abandono escolar.

Vinculación con la Deserción Escolar

Los elementos descritos —infraestructura deficiente, políticas insuficientes y condiciones socioeconómicas adversas— crean un entorno que propicia la deserción escolar en las zonas rurales. La falta de acceso a recursos educativos y las presiones económicas obligan a muchos estudiantes a priorizar la supervivencia sobre la educación, perpetuando un ciclo de desigualdad que este estudio busca analizar y abordar.

Marco Teórico

El análisis predictivo de la deserción escolar en zonas rurales de Colombia exige un sustento teórico sólido que integre las dimensiones sociales, económicas y psicológicas involucradas en este fenómeno. Para ello, este estudio se apoya en tres enfoques fundamentales: la Teoría del Capital Social de Putnam (2000), la Teoría del Conflicto Social de Marx y Engels (1848) y la Teoría del Comportamiento Planificado de Ajzen (1991). Estas teorías se han seleccionado cuidadosamente por su capacidad para explicar los factores multidimensionales que subyacen al abandono escolar, alineándose con los objetivos de la investigación de detectar patrones de riesgo y diseñar estrategias de intervención efectivas.

Teoría del Capital Social (Putnam, 2000)

La Teoría del Capital Social, propuesta por Robert Putnam, destaca la importancia de las redes sociales, la confianza mutua y las normas de reciprocidad como pilares para el fortalecimiento de las comunidades. En el marco de esta investigación, este enfoque resulta esencial para explorar cómo las dinámicas de cohesión comunitaria y el apoyo familiar pueden influir en la permanencia escolar en entornos rurales. Putnam sostiene que un capital social robusto facilita mejores resultados colectivos, lo que sugiere que las redes de apoyo sólidas podrían ser clave para reducir los riesgos de deserción. Este enfoque contribuye al objetivo de

identificar patrones al permitir el análisis de variables como la participación comunitaria y la infraestructura educativa, al tiempo que orienta hacia intervenciones que refuercen estas redes como medida preventiva.

Teoría del Conflicto Social (Marx & Engels, 1848)

Desarrollada por Karl Marx y Friedrich Engels, la Teoría del Conflicto Social examina cómo las desigualdades estructurales y las relaciones de poder perpetúan la exclusión y la marginalización. En el contexto de esta investigación, este marco teórico resulta crucial para comprender cómo factores como la pobreza y las condiciones socioeconómicas en zonas rurales pueden condicionar el acceso y la continuidad en la educación. Al abordar la deserción como un problema estructural, esta teoría permite identificar patrones de riesgo asociados a inequidades sistémicas, como la distribución desigual de recursos o las barreras derivadas de contextos de conflicto. Además, orienta la formulación de intervenciones hacia políticas que enfrenten las causas raíz de estas desigualdades, promoviendo una mayor equidad educativa.

Teoría del Comportamiento Planificado (Ajzen, 1991)

La Teoría del Comportamiento Planificado, formulada por Icek Ajzen, explica las decisiones individuales como el resultado de actitudes, normas subjetivas y el control percibido sobre las acciones. Aplicada a la deserción escolar, esta teoría ofrece un lente para analizar las motivaciones y percepciones de estudiantes y familias rurales frente a la continuidad en los estudios. Factores como las expectativas sobre la utilidad de la educación o las presiones sociales pueden influir en la decisión de permanecer o abandonar la escuela, lo que la hace idónea para detectar patrones de riesgo relacionados con el comportamiento individual y colectivo. Asimismo, sugiere intervenciones prácticas, como iniciativas de sensibilización o incentivos, que modifiquen estas percepciones y fortalezcan la intención de mantenerse en el sistema educativo.

Estas tres teorías se entrelazan de manera armónica para construir un marco integral: La Teoría del Capital Social ilumina las dinámicas comunitarias, la Teoría del Conflicto Social profundiza en las desigualdades estructurales y la Teoría del Comportamiento Planificado aborda las decisiones individuales. Juntas, proporcionan una base teórica coherente que sustenta la identificación de patrones de riesgo y la propuesta de estrategias preventivas, alineándose plenamente con los objetivos de esta investigación.

Marco Conceptual

Para este estudio, se establecen las siguientes definiciones clave:

- **Deserción Escolar:** Se define como el abandono prematuro y voluntario del sistema educativo formal antes de culminar el ciclo básico obligatorio (Gómez Triana, 2016).
- **Zonas Rurales:** Áreas geográficamente dispersas, caracterizadas por baja densidad poblacional y economías predominantemente agrícolas (DANE, 2022).
- **Factores Predictivos (MEN, 2022):** Variables cuantificables que permiten anticipar la probabilidad de deserción escolar, clasificadas en:
 - **Socioeconómicos:** ingreso familiar, nivel educativo de los padres.
 - **Educativos:** rendimiento académico, repitencia.
 - **Contextuales:** distancia al colegio, infraestructura educativa.

Marco Normativo

El marco normativo de esta investigación se fundamenta en disposiciones nacionales e internacionales que consagran el derecho a la educación y promueven la equidad, especialmente en las zonas rurales de Colombia. Estas normas no solo contextualizan las obligaciones del Estado y las políticas educativas, sino que también ofrecen un sustento legal para identificar patrones de riesgo y diseñar intervenciones que mitiguen la deserción escolar.

Normativa Nacional

- Constitución Política de Colombia (1991): La Constitución establece principios fundamentales que guían esta investigación. El Artículo 67 reconoce la educación como un derecho esencial y un servicio público, asignando al Estado la responsabilidad de garantizar su acceso y permanencia, lo que resulta clave para identificar áreas de vulnerabilidad en contextos rurales. El Artículo 68 involucra a la sociedad y la familia en el proceso educativo, sugiriendo enfoques colaborativos que pueden fortalecer las estrategias de intervención. Por su parte, el Artículo 70 promueve el desarrollo cultural y social mediante la educación, destacando la necesidad de priorizar las zonas rurales en las políticas públicas.
- Ley General de Educación (Ley 115 de 1994): Esta ley define los principios del sistema educativo colombiano, enfatizando la equidad y el acceso universal. Su enfoque en las zonas rurales permite identificar patrones de riesgo relacionados con disparidades regionales y orienta la creación de intervenciones que mejoren la cobertura y la calidad educativa en estos contextos.
- Decreto 1860 de 1994: Al reglamentar la educación formal con disposiciones específicas para áreas rurales, este decreto reconoce desafíos como la dispersión geográfica. Este enfoque normativo es vital para detectar riesgos asociados a la accesibilidad y proponer soluciones adaptadas a las particularidades de estos entornos.
- Ley 715 de 2001: Esta norma prioriza la equidad territorial en la asignación de recursos educativos, ofreciendo un marco para identificar regiones con mayores necesidades y proponer intervenciones que optimicen la distribución de fondos en zonas rurales vulnerables.
- Decreto 2888 de 2007: Al promover modelos pedagógicos flexibles para áreas rurales dispersas, este decreto subraya la importancia de adaptar la educación a contextos

específicos. Esto facilita la identificación de barreras físicas y el diseño de estrategias innovadoras, como escuelas móviles, para garantizar la continuidad escolar.

Políticas Públicas Relevantes

- "Todos a Aprender" (2012): Esta política busca mejorar la calidad educativa en zonas de bajo desempeño, incluyendo áreas rurales. Su enfoque en la equidad y el fortalecimiento docente apoya la identificación de riesgos relacionados con el rendimiento escolar y la formulación de intervenciones que eleven los estándares educativos.
- "Pacto por la Educación Rural" (2019): Comprometido con mejorar la infraestructura y el acceso en zonas rurales, este pacto proporciona un sustento para detectar patrones de riesgo ligados a condiciones materiales y proponer soluciones como transporte escolar o modernización de instalaciones.

Normativa Internacional

- Declaración Universal de Derechos Humanos (1948) – Artículo 26: Al establecer la educación como un derecho universal, esta norma refuerza la necesidad de identificar y abordar riesgos que impidan el acceso equitativo en zonas rurales, alineándose con los objetivos de intervención de la investigación.
- Convención sobre los Derechos del Niño (1989) – Artículos 28 y 29: Esta convención obliga a los Estados a garantizar la equidad educativa y reducir la deserción, especialmente entre poblaciones vulnerables como los niños rurales. Su enfoque respalda la identificación de perfiles de riesgo y el diseño de medidas específicas para proteger este derecho.
- Objetivos de Desarrollo Sostenible (ODS) – ODS 4 (Educación de Calidad): El ODS 4 promueve una educación inclusiva y equitativa para todos. Este compromiso global

orienta la investigación hacia la detección de patrones de exclusión en zonas rurales y la propuesta de estrategias que contribuyan a su cumplimiento.

Normativa sobre Datos Abiertos y Transparencia

- Ley 1712 de 2014: Al garantizar el acceso a datos públicos, esta ley facilita el análisis de información oficial para identificar patrones de riesgo, asegurando la robustez de las intervenciones propuestas.
- Decreto 103 de 2015: Este decreto fomenta la reutilización de datos abiertos, apoyando la creación de modelos predictivos replicables que refuercen las estrategias de política pública en educación rural.

Este marco normativo ofrece una estructura coherente que vincula las obligaciones legales con los objetivos de la investigación. Al destacar el derecho a la educación y la equidad, proporciona una base sólida para detectar riesgos estructurales y diseñar intervenciones prácticas que respondan a las necesidades de las zonas rurales de Colombia.

Metodología

Método

El proyecto adoptará un enfoque cuantitativo no experimental, utilizando técnicas descriptivas y predictivas para analizar patrones de deserción escolar en zonas rurales. Según Hernández, Fernández y Baptista (2018), este enfoque permite obtener una visión más completa del fenómeno al integrar distintas perspectivas analíticas, facilitando la identificación de patrones y factores subyacentes en la deserción escolar rural.

Tipo de Estudio

Este estudio se clasifica en tres categorías principales:

- **Descriptivo:** Para caracterizar los patrones y tendencias de la deserción escolar en zonas rurales de Colombia.
- **Correlacional:** Para analizar la relación entre factores socioeconómicos, educativos y geográficos con el abandono escolar.
- **Predictivo:** Para desarrollar modelos de machine learning capaces de anticipar casos de deserción con base en variables cuantificables.

Recolección de Datos

La recolección de datos se llevará a cabo en distintas fases, utilizando fuentes secundarias:

Fuentes Secundarias

- **Datos oficiales del Ministerio de Educación Nacional:** Estadísticas de matrícula, infraestructura educativa y programas de apoyo escolar.
- **Información del DANE:** Características socioeconómicas de las familias y nivel de pobreza en las regiones analizadas.

- Datos del Departamento Nacional de Planeación (DNP): Indicadores de desarrollo rural y acceso a servicios educativos.
- Estadísticas del ICFES: Resultados académicos de los estudiantes como indicador del desempeño escolar.
- Estadísticas de conflicto armado en Colombia CNMH: Cifras del conflicto armado en Colombia, como indicador de vulnerabilidad escolar.

Procesamiento de Datos

El procesamiento de los datos constituye una etapa crítica en la investigación, ya que garantiza la calidad, coherencia y utilidad de la información utilizada en los análisis descriptivos y predictivos. Este proceso se llevó a cabo mediante varias fases sistemáticas, descritas a continuación:

- Limpieza y validación inicial: Se realizó la detección y corrección de errores, inconsistencias o duplicados en las bases de datos provenientes de fuentes oficiales. Este proceso permitió asegurar que los datos fueran fiables y estuvieran libres de valores aberrantes o entradas incorrectas (Rahm & Do, 2000).
- Unificación de formatos y estructuras: Dado que los datos fueron recolectados de múltiples fuentes institucionales, fue necesario estandarizar los nombres de variables, los formatos de fecha y los tipos de datos, con el fin de permitir su integración y comparación. Esta etapa se conoce como data wrangling y es fundamental en estudios con múltiples conjuntos de datos heterogéneos (Kandel et al., 2011).
- Codificación de variables cualitativas: Las variables categóricas, como la zona geográfica o el tipo de institución educativa, fueron transformadas en variables numéricas mediante técnicas como la codificación one-hot o codificación ordinal, según correspondiera.

Esto fue necesario para facilitar su inclusión en modelos estadísticos y de aprendizaje automático (James et al., 2013).

- Tratamiento de valores atípicos y datos faltantes: Se aplicaron métodos de imputación como la imputación por la media, la moda o regresión, dependiendo del tipo de variable y el patrón de los datos faltantes. Asimismo, se identificaron y trataron valores atípicos utilizando medidas estadísticas como el rango intercuartílico (IQR) (Little & Rubin, 2019).
- Documentación detallada del proceso: Cada una de las transformaciones aplicadas a los datos fue registrada sistemáticamente, siguiendo principios de trazabilidad y reproducibilidad científica. Esto permite que el análisis pueda ser replicado por otros investigadores y garantiza la transparencia del estudio (Peng, 2011).

Instrumentos Principales.

Para llevar a cabo el procesamiento, análisis y visualización de los datos, se utilizaron diversos instrumentos y herramientas tecnológicas, seleccionadas por su eficiencia, escalabilidad y amplia adopción en el ámbito de la ciencia de datos:

- Python: Es un lenguaje de programación ampliamente utilizado en análisis de datos, modelado estadístico y aprendizaje automático. En esta investigación, se emplearon bibliotecas especializadas como pandas para la manipulación de datos, Scikit-Learn para la construcción de modelos predictivos y Matplotlib y Seaborn para la visualización básica de datos (Van Rossum & Drake, 2009; Pedregosa et al., 2011).
- Power BI: Herramienta de inteligencia de negocios desarrollada por Microsoft que permite la creación de visualizaciones interactivas, paneles y reportes automatizados. Se utilizó para presentar de manera comprensible los hallazgos a partir de gráficos dinámicos y mapas geoespaciales, facilitando la interpretación de patrones en los datos (Ferrari & Russo, 2016).

- Plataformas de procesamiento masivo de datos: Para el manejo de grandes volúmenes de datos provenientes de bases de datos gubernamentales, se recurrió a entornos de cómputo como Databricks, los cuales permiten el procesamiento distribuido a través del motor Apache Spark, mejorando significativamente el rendimiento al trabajar con datos a gran escala (Zaharia et al., 2016).
- Sistemas de gestión de bases de datos (SGBD): Se usó herramientas como MySQL para el almacenamiento estructurado de los datos. Estas bases de datos relacionales permitieron mantener la integridad de los datos y facilitar su consulta mediante lenguaje SQL, lo cual fue clave para el análisis exploratorio inicial y la preparación de datos (Elmasri & Navathe, 2016).

Matriz de Recolección de Datos.

Tabla 1

Matriz de Recolección de Datos por Variable

Variable	Definición	Fuente	Procesamiento	Rol en el Análisis
Tasa de deserción escolar	Porcentaje de estudiantes que abandonan la escuela antes de completar el ciclo	Ministerio de Educación Nacional	Limpieza (valores nulos), estandarización	Variable dependiente (outcome)
Tasa de aprobación	Porcentaje de estudiantes que aprueban el año escolar	Ministerio de Educación Nacional	Limpieza, estandarización	Variable independiente (predictor)
Tasa de repitencia	Porcentaje de estudiantes que repiten el año	Ministerio de Educación Nacional	Limpieza, estandarización	Variable independiente (predictor)

Variable	Definición	Fuente	Procesamiento	Rol en el Análisis
Tasa de reprobación	Porcentaje de estudiantes que no aprueban el año escolar	Ministerio de Educación Nacional	Limpieza, estandarización	Variable independiente (predictor)
Tasa de matrícula	Porcentaje de niños en edad escolar matriculados	Ministerio de Educación Nacional	Limpieza, estandarización	Variable independiente (predictor)
Índice de Pobreza Multidimensional (IPM)	Medida de privaciones en salud, educación, trabajo y condiciones de vida	DANE	Normalización	Variable independiente (predictor)
Casos de desaparición forzada	Número de casos reportados de desaparición forzada	Centro Nacional de Memoria Histórica (CNMH)	Transformación logarítmica	Variable independiente (predictor)
Casos de reclutamiento forzado	Número de casos reportados de reclutamiento forzado	Centro Nacional de Memoria Histórica (CNMH)	Transformación logarítmica	Variable independiente (predictor)
Casos de violencia sexual	Número de casos reportados de violencia sexual	Centro Nacional de Memoria Histórica (CNMH)	Transformación logarítmica	Variable independiente (predictor)

Fases del Proyecto

El proyecto se desarrollará en cinco fases principales, distribuidas a lo largo de 12 meses, con actividades específicas asignadas a cada mes:

Fase 1: Planificación y Revisión Bibliográfica (Meses 1-6).

Objetivo: Establecer las bases conceptuales, teóricas y metodológicas del proyecto.

- Meses 1-2: Revisión bibliográfica inicial sobre deserción escolar, análisis predictivo y educación rural.
- Meses 2-3: Formulación del planteamiento del problema basado en evidencia empírica.
- Meses 3-4: Diseño del marco teórico y conceptual, incluyendo factores socioeconómicos y el impacto del conflicto armado.
- Meses 5-6: Identificación de fuentes de datos oficiales y diseño del anteproyecto.

Fase 2: Recolección y Procesamiento de Datos (Meses 7-8).

Objetivo: Obtener, organizar y preparar los datos para su análisis.

- Mes 7: Descarga y organización de información proveniente de bases de datos oficiales (Ministerio de Educación, DANE, entre otros).
- Mes 8: Limpieza y tratamiento de datos (unificación de formatos, eliminación de valores atípicos y documentación de la base de datos).

Fase 3: Análisis Descriptivo y Predictivo (Meses 8-10).

Objetivo: Explorar patrones, correlaciones y construir modelos predictivos.

- Mes 8: Análisis descriptivo inicial para identificar tendencias y relaciones entre variables.

Resultados

En esta sección se presentan los resultados del análisis cuantitativo llevado a cabo en el marco del estudio "Análisis Predictivo de la Deserción Escolar en Áreas Rurales de Colombia: Identificación de Factores de Riesgo". Este análisis se fundamentó en datos oficiales obtenidos de diversas fuentes, tales como el Ministerio de Educación Nacional (MEN), el Departamento Administrativo Nacional de Estadística (DANE) y el Centro Nacional de Memoria Histórica (CNMH). Se integraron seis bases de datos que incluyen variables educativas, socioeconómicas, de violencia y pobreza multidimensional, las cuales fueron procesadas y analizadas mediante técnicas descriptivas, correlacionales y predictivas, utilizando herramientas como regresión lineal y árboles de decisión, implementadas en Python con bibliotecas como Pandas, Scikit-learn y Matplotlib en un entorno de Jupyter Notebook. A continuación, se detallan los hallazgos organizados en dos grandes apartados:

1. Primero, un contexto regional que describe las áreas más afectadas por indicadores educativos y factores de riesgo como violencia y pobreza multidimensional.
2. Segundo, los procedimientos y resultados específicos derivados del análisis cuantitativo. Este enfoque busca proporcionar una visión integral y fundamentada de los factores asociados a la deserción escolar en zonas rurales de Colombia.

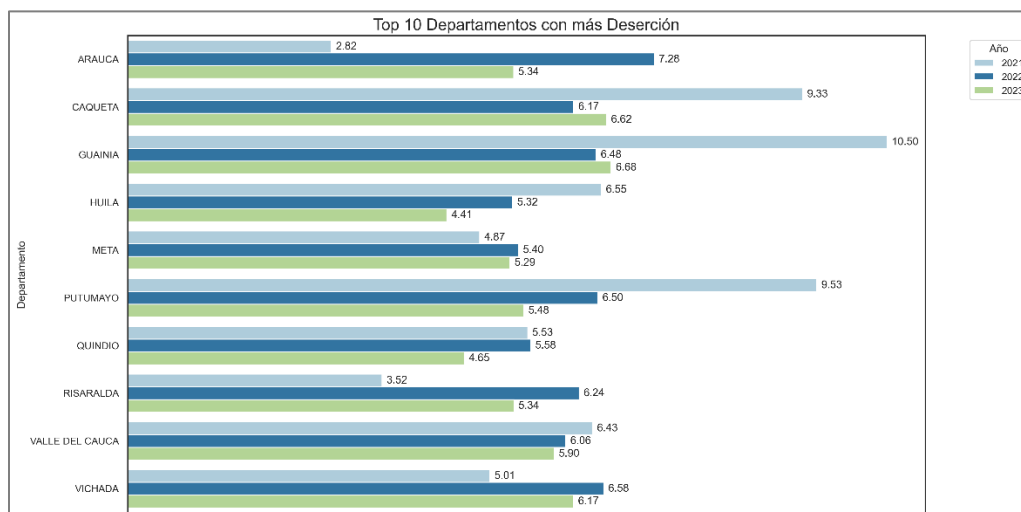
Contexto Regional: Indicadores Educativos, Violencia y Pobreza Multidimensional.

1. Departamentos y municipios con mayores índices de deserción, repitencia, matrícula y aprobación: El análisis descriptivo inicial permitió identificar las regiones con mayores desafíos educativos en zonas rurales. A continuación, se presentan los departamentos y municipios más representativos para cada indicador:

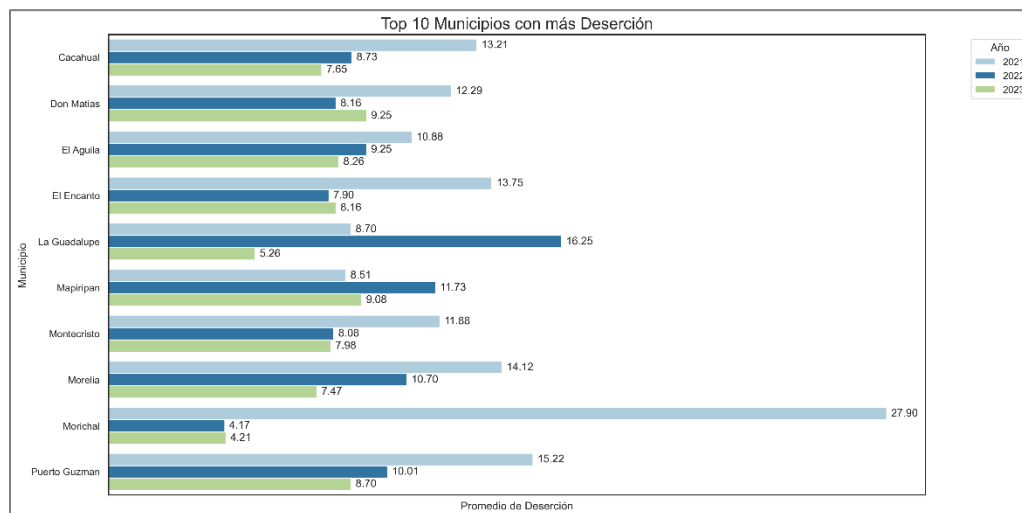
1.1. Deserción escolar: Los departamentos con las tasas más altas para el 2023 son Guainía (6.7%), Caquetá (6.6%), y Vichada (6.1%).

Figura 1

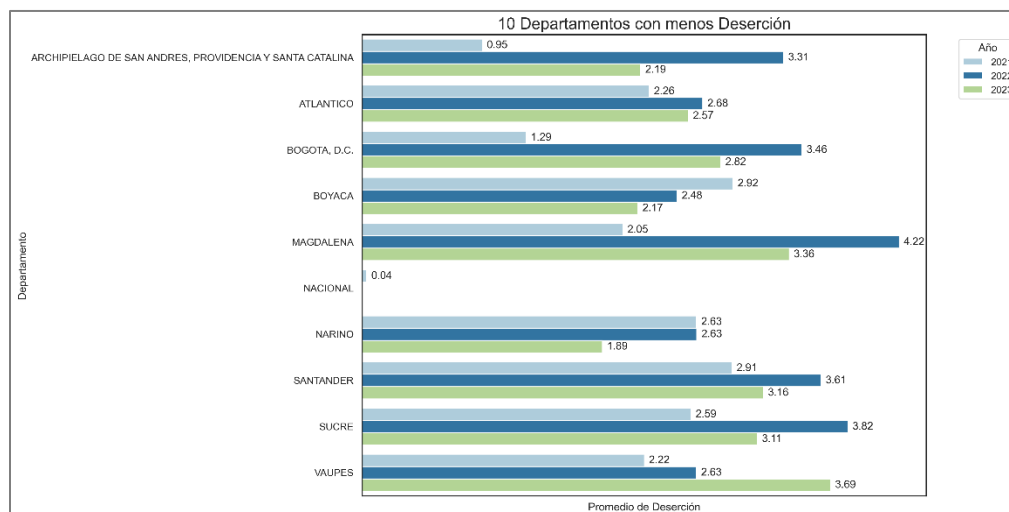
Top 10 Departamentos con Mayor Tasa de Deserción



A nivel municipal, destacan para el año 2023 Don Matías (Antioquia) con una tasa de 9.2%, Puerto Guzmán (Putumayo) con 8.7% y El Aguila (Valle del Cauca) con 8.2%.

Figura 2*Top 10 Municipios con Mayor Tasa de Deserción*

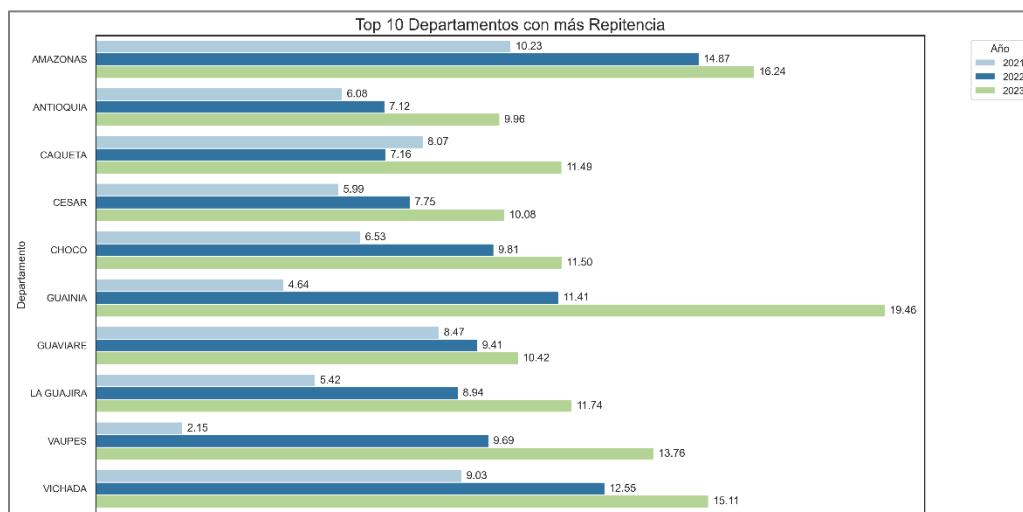
En contraste, las tasas más bajas departamentales para el 2023 se observan en Nariño (1.9%), Boyacá (2.1%) y El archipiélago de San Andrés, Providencia y Santa Catalina (2.2%).

Figura 3*Top 10 Departamentos con Menor Tasa de Deserción*

Repitencia: Los mayores índices para el 2023 se registran en Guainía (19.4%), Amazonas (16.2%) y Vichada (15.1%).

Figura 4

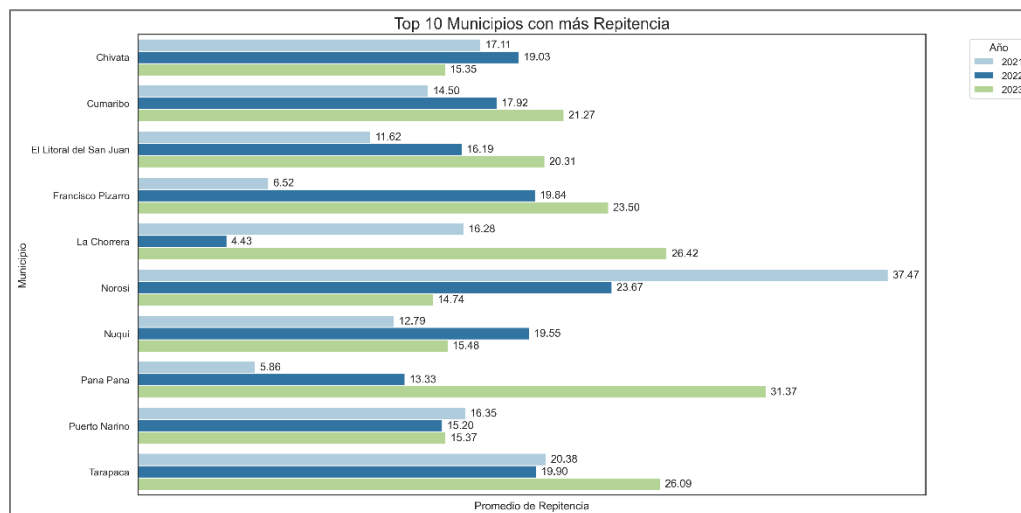
Top 10 Departamentos con Mayor Tasa de Repitencia



Entre los municipios para el 2023, Pana Pana (Guainía) alcanza un 31.3%, seguida por La Chorrera (Amazonas) con 26.4% y finalmente Tarapaca (Amazonas) con 26%.

Figura 5

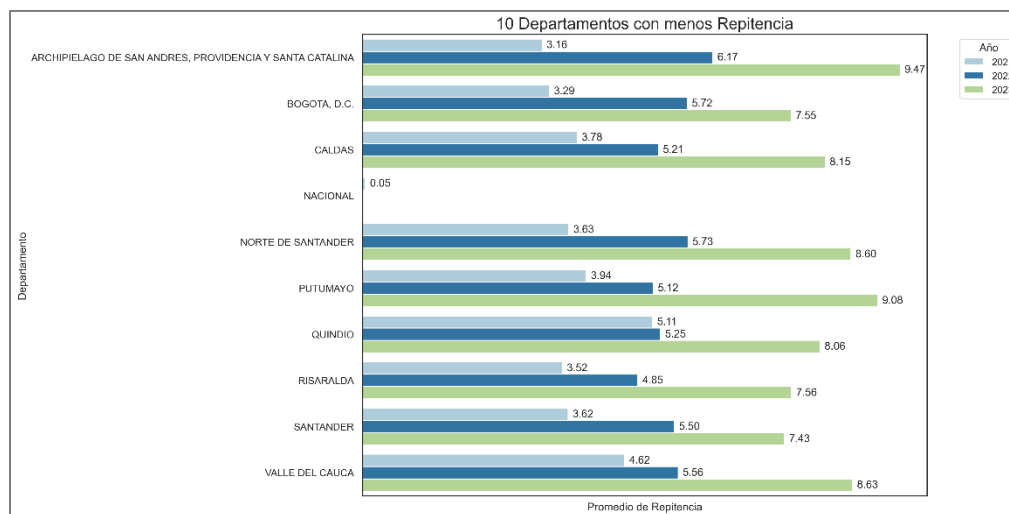
Top 10 Municipios con Mayor Tasa de Repitencia



Por su parte, las tasas más bajas para el 2023 se encuentran en Santander (7.43%), Bogotá D.C. (7.55%) y Risaralda (7.56%).

Figura 6

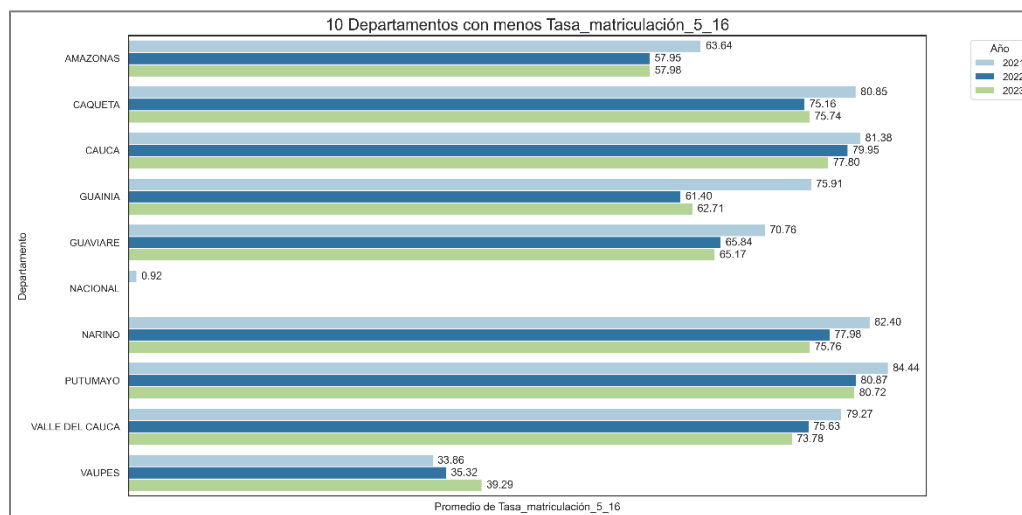
Top 10 Departamentos con Menor Tasa de Repitencia



Matrícula (5-16 años): Las tasas más bajas de matrícula para el 2023 se observan en Vaupés (39.3%), Amazonas (57.8%) y Guainía (62.7%).

Figura 7

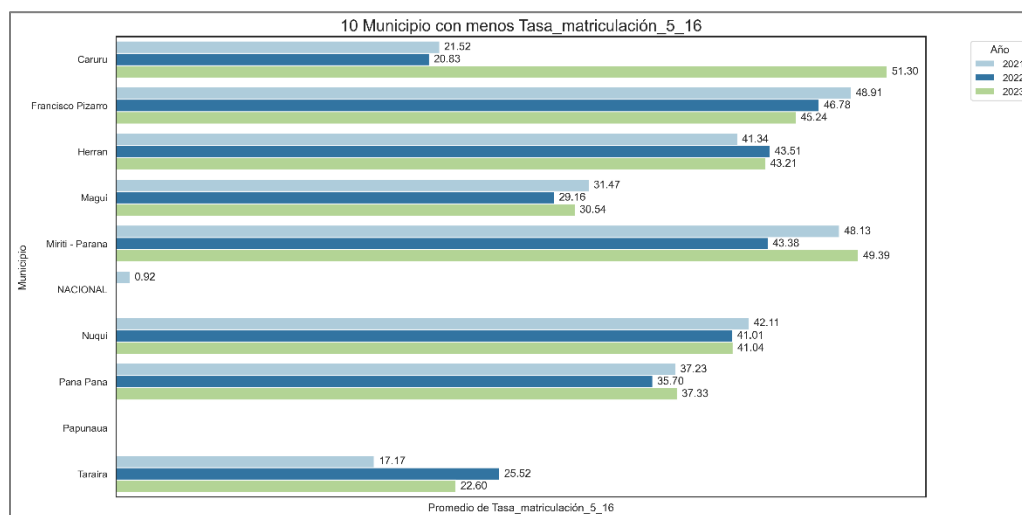
Top 10 Departamentos con Menor Tasa de Matrícula



A nivel municipal para el 2023, Taraira (Vaupés) registra un 22.6%, mientras que Magui (Nariño) tiene un 30.5%.

Figura 8

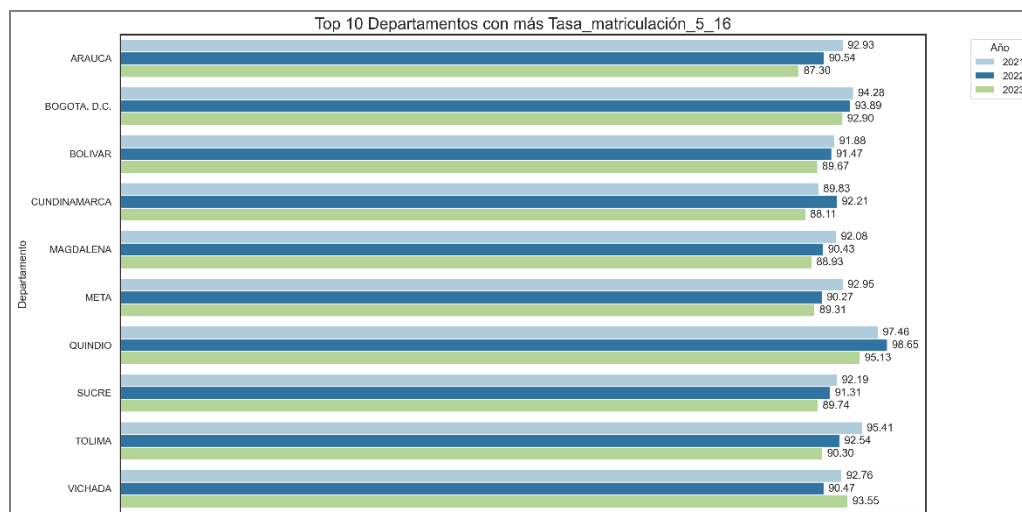
Top 10 Municipios con Menores Tasas de Matrícula



En contraste, las tasas más altas de matrícula para el 2023 se presentan en Quindío (95.1%), Vichada (93.5%) y Bogotá D.C (92.9 %).

Figura 9

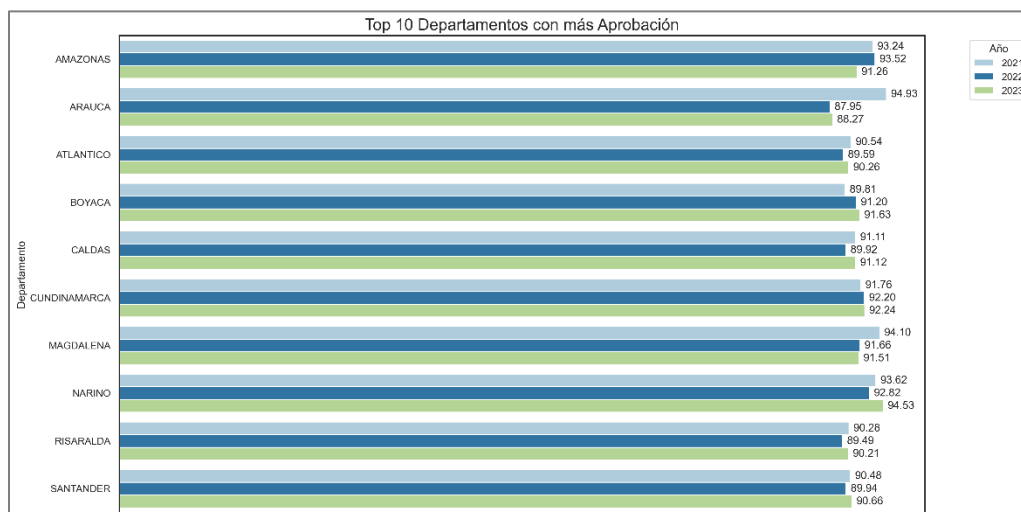
Top 10 Departamentos con Mayores Tasas de Matrícula



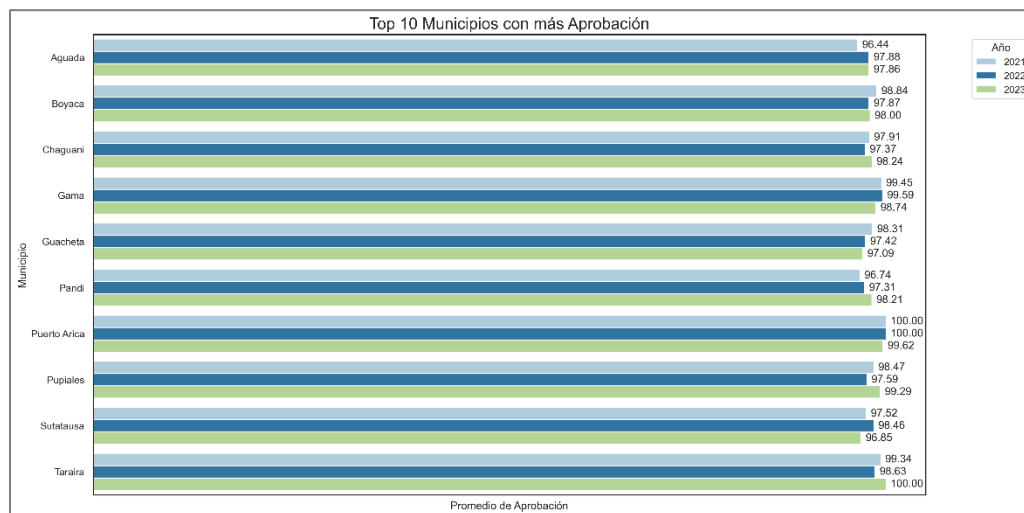
Aprobación escolar: Los departamentos con mayores tasas de aprobación para el 2023 son Nariño (94.5%), Cundinamarca (92.2%) y Amazonas (91.2%).

Figura 10

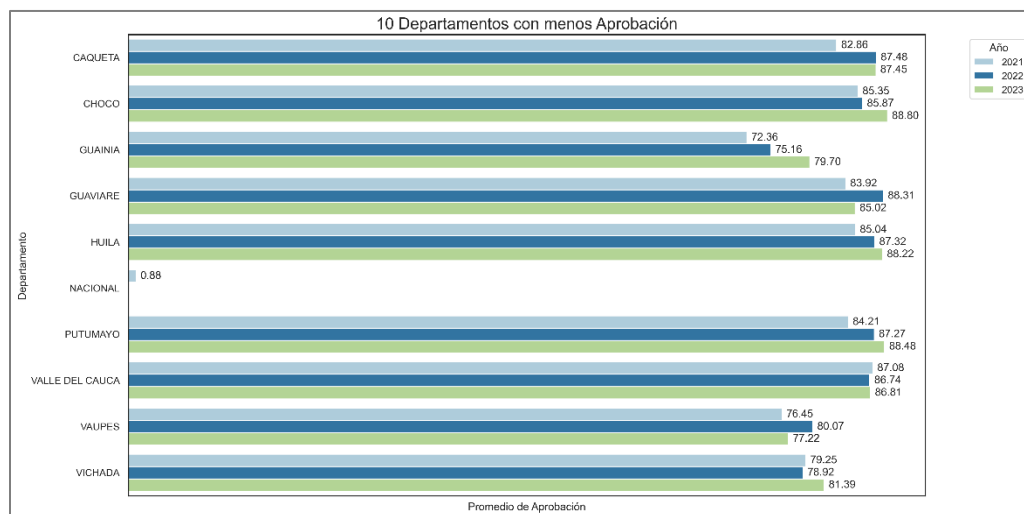
Top 10 Departamentos con Mayor Tasa de Aprobación



Entre los municipios para el 2023, Taraira (Vaupés) alcanza un 100% y Pupiales (Nariño) un 99.2%.

Figura 11*Top 10 Municipios con Mayor Tasa de Aprobación*

Por el contrario, las tasas más bajas por departamento para el 2023 se encuentran en Vaupés (77.2%), Guainía (79.7%) y Vichada (81.3%).

Figura 12*Top 10 Departamentos con Menor Tasa de Aprobación*

Estos datos evidencian una clara disparidad regional, con departamentos y municipios de la periferia del país mostrando indicadores educativos más desfavorables en comparación con las áreas urbanas y centrales.

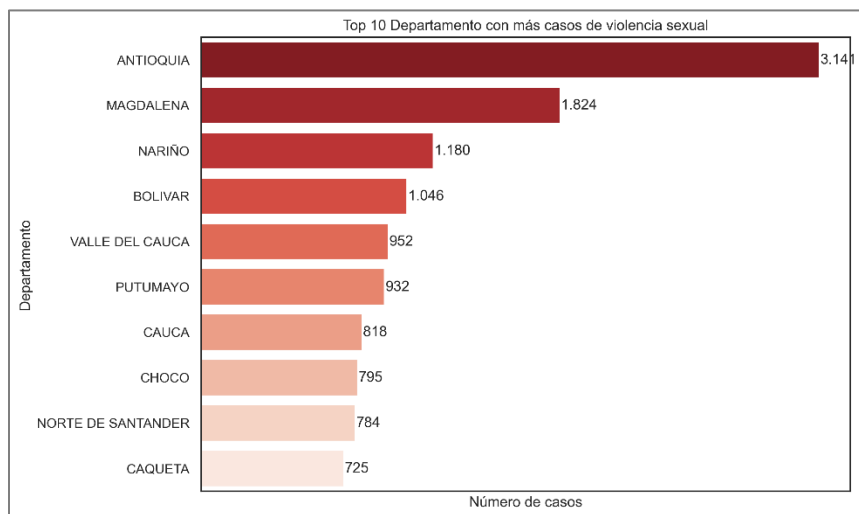
Departamentos y Municipios con Mayores Casos de Violencia

El estudio incorporó variables de violencia que podrían influir en los resultados educativos, basándose en datos del CNMH. Las cifras más relevantes son:

- **Violencia sexual:** Los departamentos con mayor número de casos reportados son Antioquia (3141 casos), Magdalena (1824 casos) y Nariño (1180 casos). A nivel municipal, Medellín (Antioquia) registra 450 casos, Cali (Valle del Cauca) 380 casos y Soacha (Cundinamarca) 320 casos.

Figura 13

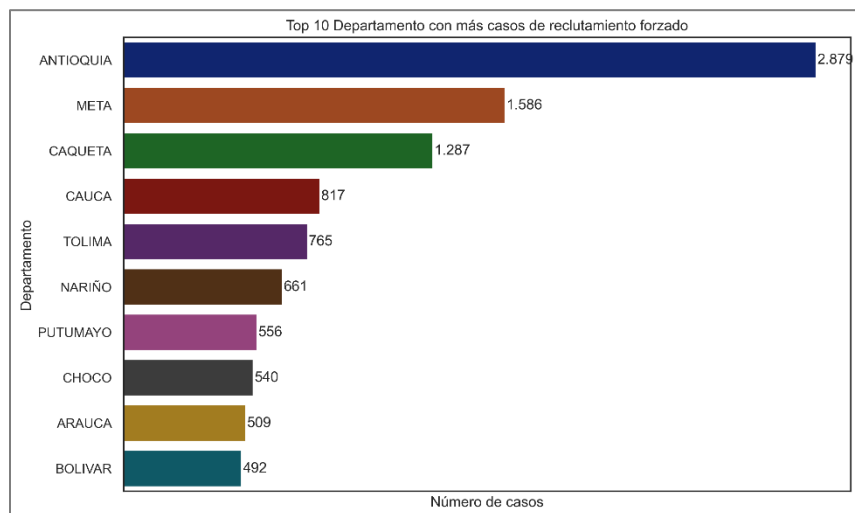
Top 10 Departamentos con más Casos de Violencia Sexual



- **Reclutamiento forzado:** Las cifras más altas se presentan en Antioquia (2879 casos), Meta (1586 casos) y Caquetá (1287 casos).

Figura 14

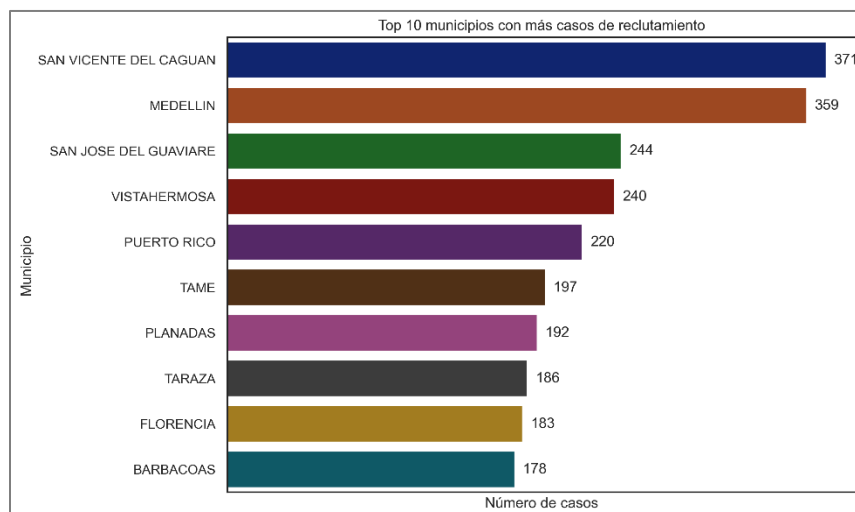
Top 10 Departamentos con más Casos de Reclutamiento Forzado



Entre los municipios, San Vicente del Caguán (Caquetá) reporta 371 casos, Medellín (Antioquia) 359 casos y San José del Guaviare (Guaviare) 244 casos.

Figura 15

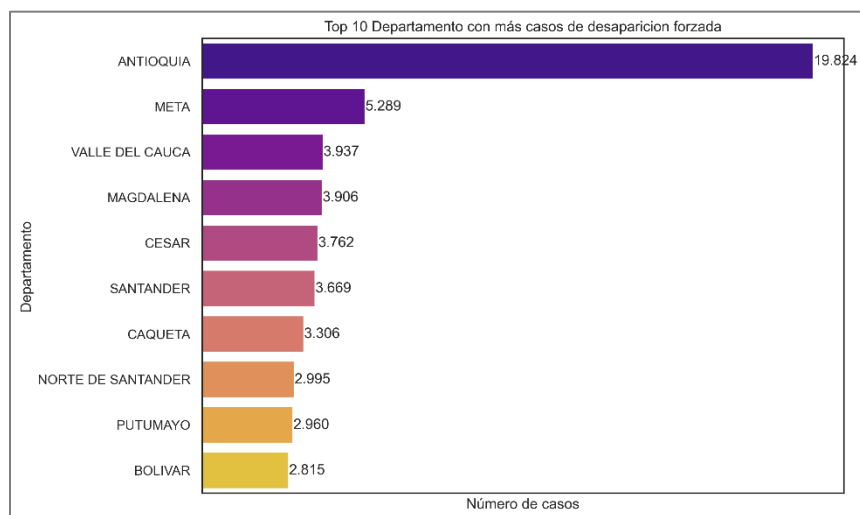
Top 10 Municipios con Mayores Casos de Reclutamiento



- Desaparición forzada: Los departamentos más afectados son Meta (420 casos), Cauca (380 casos) y Putumayo (350 casos).

Figura 16

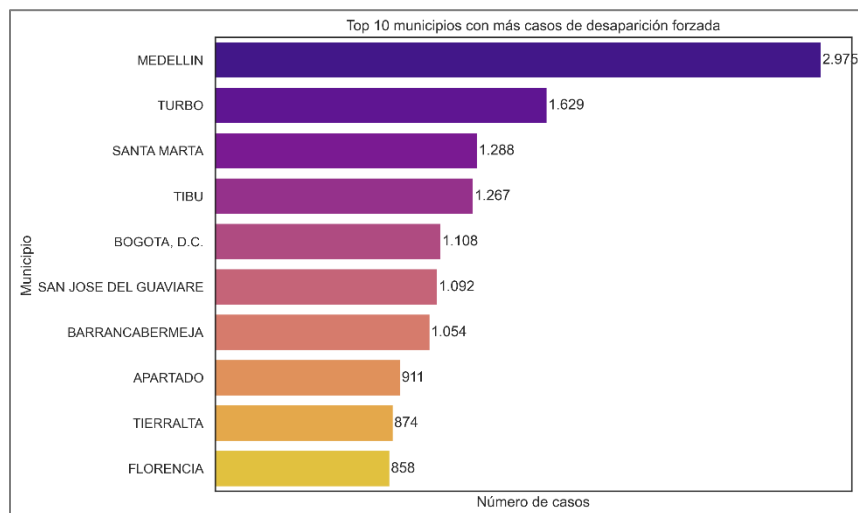
Top 10 Departamentos con más Casos de Desaparición Forzada



A nivel municipal, Medellín (Antioquia) registra 2975 casos y Turbo (Antioquia) 1629 casos reportan las mayores cifras.

Figura 17

Top 10 Municipios con más Casos de Desaparición Forzada



Estas cifras reflejan la prevalencia de la violencia y de conflicto armado tanto en zonas rurales como cabeceras municipales de Colombia.

Índice de Pobreza Multidimensional (IPM)

El Índice de Pobreza Multidimensional (IPM), calculado por el DANE, mide privaciones en salud, educación, trabajo y condiciones de vida. Este índice permite identificar de manera integral las carencias que afectan la calidad de vida de la población en los diferentes territorios del país.

Según la encuesta IPM 2024, los departamentos con los mayores índices son:

- Vichada: 70.2%
- Guainía: 49%
- La Guajira: 39.3%

En contraste, los departamentos con los índices más bajos son:

- Bogotá D.C: 5.4%

- San Andrés: 6%
- Valle del Cauca: 6.2%

La coincidencia entre altos niveles de pobreza multidimensional y tasas elevadas de deserción escolar sugiere una relación significativa entre estas variables, lo cual es un aspecto clave para el análisis de políticas públicas.

Estadísticas Descriptivas por Variable para el Año 2024

Se analizaron las principales variables del IPM en Colombia, agrupadas por total nacional, cabeceras municipales y zonas rurales. Se calcularon medidas resumen (media, máximo, mínimo y suma) para cada variable.

Resultados clave:

- Trabajo informal: Presenta los valores más altos, con un promedio nacional total de 79.5%, que se incrementa notablemente en zonas rurales (media rural: 88.9%). Esto refleja la vulnerabilidad laboral en las áreas no urbanas.
- Bajo logro educativo: es la segunda variable con valores significativos, con una media total nacional de 46.9%, notablemente mayor en zonas rurales (69%) que en cabeceras (35.4%).
- Variables relacionadas con condiciones de vida, como sin acceso a fuente de agua mejorada y material inadecuado de pisos, muestran promedios rurales más del doble que en las cabeceras, indicando brechas de infraestructura básicas importantes.
- Otras dimensiones como analfabetismo (9.9% total, 15.7% rural), desempleo de larga duración (12.7%) y hacinamiento crítico (7.8%) reflejan condiciones que, aunque menos prevalentes, afectan significativamente ciertas regiones.

En síntesis, las condiciones socioeconómicas y de infraestructura son consistentemente más precarias en zonas rurales frente a cabeceras, evidenciando desigualdad territorial en los indicadores del IPM.

Brechas Rurales contra Cabeceras Municipales

Se calculó la brecha entre las tasas rurales y urbanas (cabeceras municipales) para cada variable por departamento, identificando las disparidades más marcadas.

Resultados más destacados:

- Analfabetismo: presenta las brechas absolutas más altas, con departamentos como Vichada (31.3 puntos porcentuales), La Guajira (25.6) y Chocó (20.4), mostrando mayor rezago rural.

- En bajo logro educativo, las brechas también son muy grandes: Vichada con 48.2 puntos, Caldas con 44.5, y Norte de Santander con 41.5, revelando una mayor desigualdad educativa entre zonas urbanas y rurales.

Otras variables con brechas significativas incluyen:

- Inadecuada eliminación de excretas, con Vaupés con una diferencia de 72.3 puntos.

- Material inadecuado de pisos, con Santander, Vichada y La Guajira presentando brechas mayores a 60 puntos.

- Sin acceso a fuente de agua mejorada, con Vaupés, Vichada y Caquetá superando diferencias de 70 puntos porcentuales.

Estas brechas reflejan desigualdades estructurales y acceso diferenciado a servicios básicos, educación y condiciones de vida dignas entre zonas rurales y urbanas en múltiples departamentos.

Implicaciones para Políticas Públicas y Conclusiones

Los resultados evidencian que:

- Las zonas rurales están en desventaja significativa en casi todas las dimensiones del IPM, lo cual demanda políticas focalizadas para cerrar estas brechas estructurales.
- La educación y el acceso a servicios básicos (agua, saneamiento, vivienda) son áreas críticas con enormes desigualdades territoriales.
- Departamentos como Vichada, La Guajira, Chocó y Vaupés requieren atención prioritaria debido a sus altos índices y brechas en variables claves.
- Intervenciones multisectoriales integradas que involucren educación, infraestructura y desarrollo económico rural son necesarias para reducir la pobreza multidimensional.

Este diagnóstico robusto basado en datos desagregados por año y territorio aporta evidencia cuantitativa para orientar decisiones de política pública que contribuyan a la equidad social en Colombia.

Procedimientos de Análisis

El análisis cuantitativo se desarrolló en varias etapas para garantizar la robustez de los resultados, empleando herramientas y metodologías actuales en ciencia de datos:

Integración y Limpieza de Datos

Se consolidaron las bases de datos provenientes de fuentes oficiales, abarcando indicadores de matrícula, infraestructura educativa, programas de alimentación escolar, violencia y pobreza multidimensional. La integración se realizó utilizando la biblioteca Pandas de Python (McKinney, 2010), fusionando las fuentes mediante claves comunes, como los códigos de municipios. La limpieza de datos siguió las metodologías descritas por Rahm y Do (2000),

incluyendo la eliminación de valores nulos, la estandarización de formatos y la corrección de valores atípicos. Las variables categóricas, como la zona geográfica, se codificaron numéricamente mediante one-hot encoding, un procedimiento estándar en el preprocesamiento de datos (James et al., 2013).

Análisis Descriptivo

Se calcularon estadísticas descriptivas, como media, mediana y desviación estándar, para explorar la distribución de las variables relacionadas con la deserción escolar. Se generaron visualizaciones exploratorias, incluyendo histogramas y gráficos de dispersión, siguiendo las técnicas de análisis exploratorio de datos presentadas por James et al. (2013). Estas herramientas permitieron identificar patrones y posibles asociaciones entre variables predictoras y la deserción.

Análisis de Correlación

Se emplearon coeficientes de correlación de Pearson para evaluar las relaciones lineales entre la deserción escolar y variables predictoras, como indicadores educativos, socioeconómicos y de violencia. Este análisis se basó en los métodos descritos por James et al. (2013), quienes destacan la importancia de interpretar estas relaciones en contextos multivariados. Los resultados se visualizaron mediante mapas de calor para facilitar la identificación de patrones.

Modelos de Regresión Lineal

Se implementaron modelos de regresión lineal simple y múltiple utilizando la biblioteca scikit-learn (Pedregosa et al., 2011) para analizar el impacto de variables como el nivel de conflicto y violencia en la deserción, reprobación, repitencia y tasa de matrícula. Los modelos se evaluaron según las directrices de James et al. (2013), considerando la significancia estadística de los coeficientes, el coeficiente de determinación (R^2) y el error cuadrático medio (MSE).

Árboles de Decisión

Se entrenaron modelos de árboles de decisión utilizando el algoritmo CART (Classification and Regression Trees), implementado en scikit-learn (Pedregosa et al., 2011). Se evaluó el desempeño de 19 variables predictoras de forma individual, siguiendo las metodologías expuestas por James et al. (2013), quienes destacan la capacidad de los árboles de decisión para capturar relaciones no lineales.

Visualización de Resultados

Las visualizaciones se diseñaron siguiendo los principios de Wickham (2010), quien propone un enfoque estructurado para la representación gráfica de datos. Se utilizaron gráficos de dispersión para mostrar relaciones bivariadas, histogramas de errores para evaluar la distribución de residuos y mapas de calor para resumir matrices de correlación, asegurando claridad y accesibilidad en la presentación de los hallazgos.

Hallazgos Clave

1. **Análisis Descriptivo:** El análisis descriptivo destacó la variabilidad en los indicadores educativos y socioeconómicos.
 - La tasa media de deserción media es de 3.49% con una desviación estándar de 2.17%, oscilando entre 0.67 % (archipiélago de San Andrés, Providencia y Santa Catalina) y 4.57% (Guainía).
 - La tasa de matrícula promedio es de 84.97%, con extremos en 56.9% (Guaviare) y 96.41% (Bogotá D.C.).
 - El IPM promedio nacional para el 2024 es del 17.33%, pero alcanza niveles críticos en Vichada (70.2%) y presenta menores índice en Bogotá D.C (5.4%).

2. Análisis de Correlación: Los coeficientes de correlación revelaron las siguientes relaciones entre las variables educativas y factores de conflicto armado.

El análisis de correlación indica que la variable *nivel de conflicto por desaparición forzada* presenta asociaciones muy débiles con las variables educativas consideradas. Aunque las correlaciones no son estadísticamente significativas, se observan ligeras relaciones negativas con indicadores de desempeño académico.

Las correlaciones más destacadas son:

- Deserción escolar: 0,028 (relación positiva muy débil).
- Tasa de matriculación (5 a 16 años): 0,015 (sin relación significativa).
- Repitencia (total, primaria y secundaria): entre -0,017 y -0,023 (relaciones negativas muy débiles).
- Reprobación (total, primaria y secundaria): entre -0,028 y -0,035 (relaciones inversas de baja magnitud).

Los resultados sugieren que un mayor nivel de desapariciones forzadas no se asocia significativamente con la deserción ni con otros indicadores de desempeño escolar. Sin embargo, las correlaciones negativas, aunque débiles, podrían indicar una ligera tendencia hacia una menor persistencia escolar en contextos con mayor exposición a este tipo de violencia.

En contraste con la variable anterior, el nivel de conflicto por reclutamiento forzado muestra correlaciones positivas más marcadas, particularmente con los indicadores de repitencia y reprobación escolar:

- Repitencia en secundaria: 0,105
- Repitencia total: 0,094
- Repitencia en primaria: 0,083

- Reprobación total: 0,058
- Reprobación secundaria: 0,068
- Reprobación primaria: 0,041
- Tasa de matriculación: 0,001 (sin relación significativa).
- Deserción escolar: -0,005 (relación negativa despreciable).

Estos resultados evidencian una débil pero consistente asociación entre el reclutamiento forzado y el bajo rendimiento escolar, expresado en mayores tasas de repitencia y reprobación. Esto podría estar vinculado a factores como estrés, desplazamiento, interrupción del ciclo escolar o afectaciones psicosociales de los menores.

Las correlaciones con la variable *nivel de conflicto por violencia sexual* son negativas en todos los casos, aunque de magnitud baja:

- Repitencia total: -0,052
- Repitencia secundaria: -0,044
- Repitencia primaria: -0,055
- Reprobación total, primaria y secundaria: entre -0,000 y -0,007
- Deserción escolar: -0,006
- Tasa de matriculación: -0,012

Tabla 3*Resumen de Correlaciones entre Variables de Conflicto y Académicas*

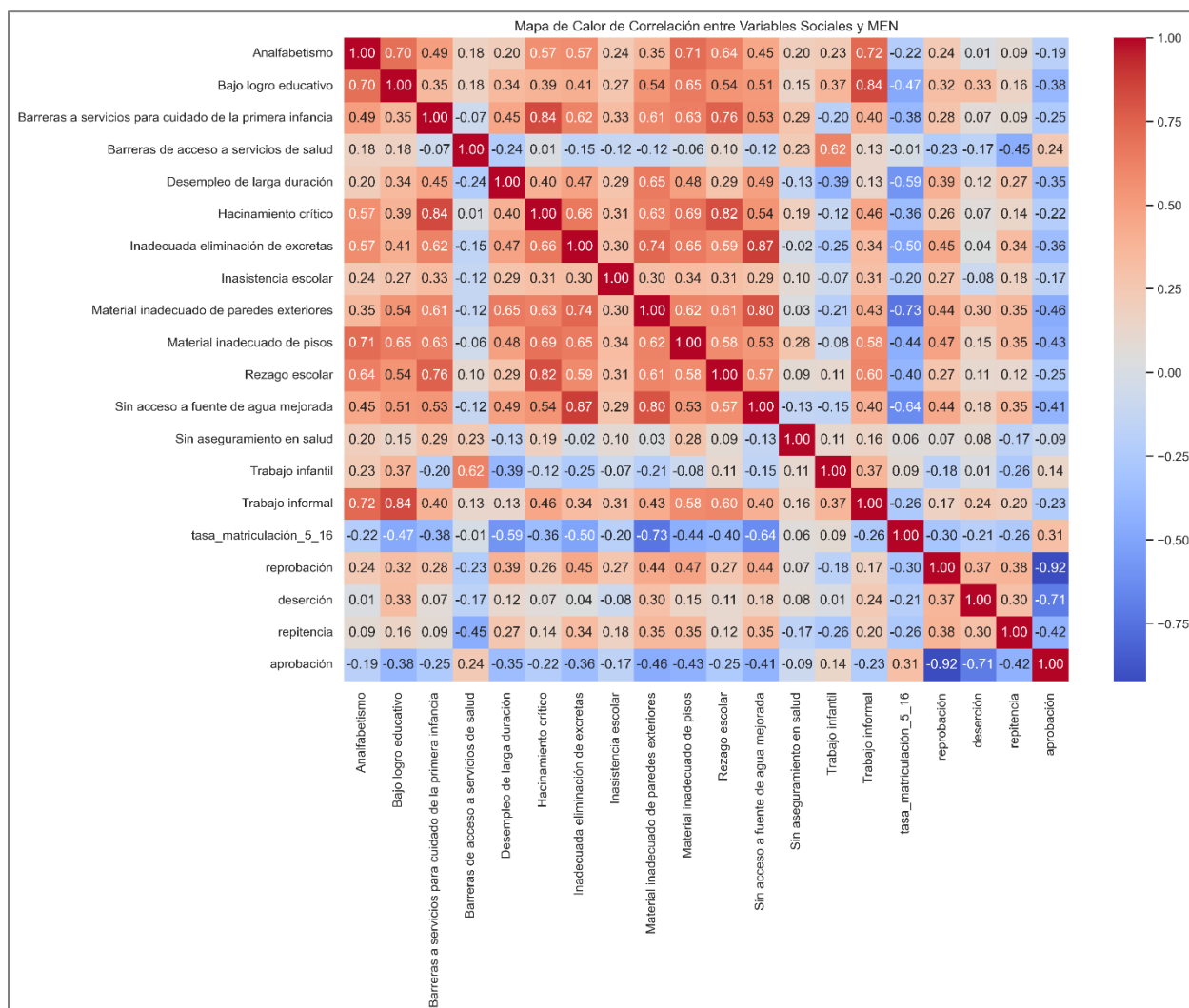
Variable Educativa	Desaparición Forzada	Reclutamiento Forzado	Violencia Sexual
Reprobación	-0,035	0,058	-0,000
Deserción escolar	0,028	-0,005	-0,006
Repitencia	-0,022	0,094	-0,052
Tasa de matriculación	0,015	0,001	-0,012

Los tipos de conflicto armado analizados muestran relaciones débiles con los indicadores educativos. No obstante, el reclutamiento forzado presenta una asociación más clara con problemas de desempeño escolar, lo que podría ser considerado en políticas educativas diferenciales para contextos rurales afectados por el conflicto.

Por su parte, los coeficientes de correlación revelaron las siguientes relaciones para las variables del IPM contra las variables académicas del MEN:

Figura 18

Mapa de Calor de Correlaciones entre Variables



Correlaciones negativas fuertes (Inversas).

- Aprobación - Reprobación (-0.92): Como es esperable, una alta aprobación se relaciona con una baja reprobación. Es la correlación más fuerte del análisis.
- Aprobación - Reptencia (-0.71): Una mayor aprobación también se asocia con menor reptencia.

- Tasa de Matriculación - Reprobación (-0.60): A mayor tasa de matrícula, se observa menor reprobación.
- Tasa de Matriculación - Repitencia (-0.53): Aumentos en la matrícula se relacionan con menos repitencia.
- Tasa de Matriculación - Deserción (-0.31): Relación inversa moderada; mayor matrícula, menor deserción.

Correlaciones positivas fuertes.

- Analfabetismo - Bajo Logro Educativo (0.70): Coherente, ya que ambas variables representan formas de exclusión educativa.
- Rezago Escolar - Reprobación (0.38): A mayor rezago, mayor probabilidad de reprobación.
- Trabajo Infantil - Deserción (0.34): El trabajo infantil se asocia con mayor deserción escolar.
- Trabajo Informal - Bajo Logro Educativo (0.72): Alta relación entre un entorno socioeconómico precario y bajo desempeño académico.

Tabla 4

Variables Educativas y su Relación con Condiciones Sociales

Variable Educativa	Correlaciones Sociales Destacadas
Reprobación	↑ Rezago escolar (0.38), ↑ Trabajo infantil (0.31), ↓ Tasa de matrícula (-0.60), ↓ Aprobación (-0.92)
Deserción	↑ Trabajo infantil (0.34), ↑ Inasistencia escolar (0.33), ↓ Tasa de matrícula (-0.31)

Variable	Correlaciones Sociales Destacadas
Educativa	
Repitencia	↑ Trabajo infantil (0.32), ↑ Trabajo informal (0.30), ↓ Tasa de matrícula (-0.53), ↓ Aprobación (-0.71)
Tasa De Matrícula	↓ Trabajo infantil (-0.50), ↓ Rezago escolar (-0.44), ↓ Reprobación (-0.60), ↓ Repitencia (-0.53)

3. Modelos de Regresión Lineal: Los modelos de regresión lineal arrojaron los siguientes resultados en los análisis de regresión entre variables educativas y tipos de conflicto armado.

- Conflicto por Desaparición Forzada

Tabla 5

Modelo de Regresión Lineal – Desaparición Forzada

Variable Educativa	Coefficiente	R ²	MSE	Interpretación
Deserción	+0.0677	-0.0080	4.71	Relación positiva muy débil; el conflicto no explica la deserción.
Reprobación	-0.1339	-0.0056	17.05	Relación negativa contraintuitiva; sin poder explicativo.
Tasa de matriculación	+0.0788	0.00028	406.34	Relación positiva débil; el modelo no tiene capacidad explicativa.
Repitencia	-0.1464	-0.0250	11.61	Relación negativa leve; el modelo es estadísticamente no significativo.

Por lo anterior, se deduce que los modelos de regresión lineal simple con el nivel de conflicto por desaparición forzada como predictor no presentan valor explicativo relevante. Los

coeficientes son pequeños, los valores de R^2 son cercanos a cero o negativos, y los errores medios son altos. Por tanto, no se evidencia un impacto estadísticamente significativo sobre las variables educativas analizadas.

- Conflicto por Reclutamiento Forzado

Tabla 6

Modelo de Regresión Lineal – Reclutamiento Forzado

Variable Educativa	Coefficiente	R^2	MSE	Interpretación
Deserción	+0.0011	0.00026	4.68	Efecto prácticamente nulo; el conflicto no predice la deserción.
Reprobación	+0.0070	-0.0004	16.96	Relación despreciable; modelo sin poder explicativo.
Tasa de matriculación	-0.0122	-0.0007	406.75	Leve disminución no significativa; modelo estadísticamente inadecuado.
Repitencia	+0.0623	-0.2340	13.97	Ligera relación positiva, pero con un modelo con muy bajo ajuste y gran error.

De igual manera el conflicto por reclutamiento forzado no muestra evidencia de asociación significativa con las variables educativas. Los modelos presentan coeficientes cercanos a cero, valores de R^2 negativos y errores altos, lo cual indica una pobre capacidad predictiva.

- Casos de violencia sexual

Tabla 7*Modelo de Regresión Lineal – Violencia Sexual*

Variable educativa	Coefficiente	R ²	MSE	Interpretación
Deserción	-0.0025	-0.0002	4.68	Relación negativa muy leve y no significativa; el modelo carece de valor explicativo.
Reprobación	-0.1293	-0.0055	17.05	Relación negativa débil y contraintuitiva; modelo sin validez predictiva.
Tasa de matriculación	-0.1530	-0.0013	407.00	Leve caída coherente en sentido, pero sin significancia estadística.
Repitencia	-0.0996	-0.0060	11.39	Disminución leve, sin respaldo estadístico ni interpretativo claro.

Asimismo, el nivel de conflicto asociado a violencia sexual tampoco presenta relación significativa con los indicadores educativos. Los modelos generados son débiles, con valores de R² negativos o cercanos a cero y errores elevados, lo que indica una ausencia de capacidad predictiva.

Por otro lado, el modelo de regresión lineal múltiple analizado desde las variables contenidas en el IPM, presenta un coeficiente de determinación R² = 1.0000, lo que implica que el 100% de la variabilidad observada en la variable dependiente (deserción escolar) es explicada por el conjunto de variables independientes incluidas en el modelo. Aunque este resultado puede parecer ideal, es poco habitual en contextos reales y puede ser indicativo de un problema de sobreajuste (overfitting), es decir, el modelo podría estar capturando de manera excesiva las características específicas de los datos de entrenamiento sin generalizar adecuadamente a datos nuevos.

El Error Cuadrático Medio (MSE) obtenido es de 0.0000, lo que indica que las predicciones del modelo son idénticas a los valores reales en el conjunto de datos utilizado. Este resultado refuerza la sospecha de sobreajuste, ya que en contextos aplicados es improbable alcanzar una predicción perfecta sin errores.

El análisis de los coeficientes permite identificar la influencia relativa de cada variable independiente sobre la deserción escolar, así como la dirección de dicha influencia.

Variables con mayor peso en el modelo:

- Reprobación escolar (-0.9997) y Aprobación escolar (-0.9997): Ambas variables presentan coeficientes muy cercanos a -1, lo que indica una fuerte relación negativa con la deserción escolar. Esto sugiere que a medida que aumentan las tasas de aprobación o reprobación, la tasa de deserción disminuye, lo cual resulta contraintuitivo en el caso de la reprobación. Este comportamiento podría explicarse por una colinealidad elevada entre las variables o por definiciones que requieren revisión en el conjunto de datos.

- Variables con influencia baja o nula: La mayoría de las demás variables del modelo, como trabajo infantil, rezago escolar, analfabetismo, entre otras, presentan coeficientes cercanos a cero. Esto indica que, bajo la estructura actual del modelo, estas variables tienen una contribución poco significativa a la predicción de la deserción escolar. No obstante, esto no implica que no tengan relevancia en la realidad, sino que la forma en que fueron medidas o incorporadas en el modelo no permite capturar su efecto adecuadamente.

4. Árboles de Decisión: Con el objetivo de evaluar la capacidad predictiva de distintas formas de violencia asociadas al conflicto armado sobre variables educativas, se implementaron modelos de árbol de decisión univariados. Las variables predictoras fueron:

desaparición forzada, reclutamiento forzado y violencia sexual. Como variables objetivo se consideraron: deserción escolar, reprobación, tasa de matrícula y repitencia.

- Desaparición forzada: Los árboles de decisión construidos con desaparición forzada como predictor presentaron resultados deficientes en todos los casos. El valor de R^2 fue cercano a cero o incluso negativo, lo que indica que el modelo no logra explicar la variabilidad de las variables educativas mejor que una predicción basada en la media. Además, los errores cuadráticos medios (MSE) fueron moderados o altos, y los coeficientes de los nodos fueron despreciables. Se concluye que no existe evidencia de relación significativa entre desaparición forzada y los indicadores educativos considerados.

- Reclutamiento forzado: Los resultados para los modelos de árbol que emplearon reclutamiento forzado como variable explicativa fueron esencialmente equivalentes a los de desaparición forzada. En todos los casos, el valor de R^2 fue nulo o negativo, y los niveles de error indicaron un bajo ajuste del modelo. Por lo tanto, el reclutamiento forzado no mostró capacidad predictiva sobre deserción escolar, reprobación, matrícula ni repitencia.

- Violencia sexual: En contraste, los árboles de decisión que utilizaron violencia sexual como variable independiente presentaron una leve mejora en el ajuste del modelo, específicamente en el caso de la reprobación escolar ($R^2 = 0.0748$). Si bien este valor sigue siendo bajo, representa la mayor capacidad explicativa obtenida entre los tres tipos de conflicto analizados. Para las demás variables educativas, los modelos mantuvieron valores de R^2 cercanos a cero y errores elevados, por lo que no se identificó una relación significativa.

- Pobreza Multidimensional (IPM): sobre los principales indicadores educativos en contextos rurales, se construyó un modelo de árbol de decisión para cada variable dependiente: deserción escolar, reprobación, tasa de matrícula y repitencia. A diferencia de los modelos

anteriores basados en variables de conflicto armado, el IPM mostró un desempeño notablemente superior en términos de capacidad predictiva.

- La aprobación escolar fue la variable mejor explicada por el IPM, con un R^2 de 0.6508 y un error cuadrático medio (MSE) de 1.48. Este valor indica que aproximadamente el 65% de la varianza en la aprobación puede explicarse por el nivel de pobreza multidimensional, lo cual representa un ajuste considerablemente fuerte en contextos sociales.

- Material inadecuado en paredes de la vivienda ($R^2 = 0.1080$)
- Barreras de salud ($R^2 = 0.0543$)
- Reprobación escolar ($R^2 = 0.0579$)
- Ausencia de acceso a agua mejorada ($R^2 = 0.0534$)
- Bajo logro educativo del hogar ($R^2 = 0.0459$)

Estos resultados, aunque de menor magnitud que la aprobación, evidencian que ciertas dimensiones del IPM tienen una relación, aunque débil, con la situación educativa de los niños en zonas rurales.

- Variables con mal desempeño predictivo: En contraste, varias dimensiones del IPM mostraron valores negativos de R^2 , lo que significa que el modelo fue menos efectivo que una predicción basada en la media. Entre ellas destacan:

- Sin aseguramiento en salud ($R^2 = -0.6179$)
- Inasistencia escolar ($R^2 = -0.2739$)
- Tasa de matrícula ($R^2 = -0.2401$)
- Condiciones como hacinamiento, barreras de primera infancia, trabajo infantil y

eliminación de excretas también presentaron valores cercanos a $R^2 = -0.20$, con errores MSE elevados.

Esto sugiere que, si bien el IPM global es útil como predictor, no todas sus dimensiones individuales poseen valor explicativo significativo en modelos univariados.

5. Visualizaciones: Las visualizaciones respaldaron los análisis cuantitativos.
 - Gráficos de líneas mostraron una relación lineal fuerte entre la aprobación y la deserción.
 - Histogramas de errores indicaron una distribución aleatoria en modelos bien ajustados.
 - Mapas de calor resaltaron correlaciones entre variables educativas y socioeconómicas, como IPM y matrícula.

Interpretación y Discusión

Los resultados de esta investigación arrojan luz sobre las dinámicas complejas de la deserción escolar en las zonas rurales de Colombia, revelando una interacción multifacética de factores educativos, socioeconómicos y contextuales que influyen en la permanencia estudiantil. Un hallazgo central es la fuerte correlación negativa entre la aprobación escolar y la deserción, con un coeficiente de -0.92 respecto a la reprobación. Esto indica que los estudiantes que logran un buen desempeño académico tienen una probabilidad significativamente menor de abandonar sus estudios. Desde la perspectiva de la Teoría del Comportamiento Planificado de Ajzen (1991), este vínculo sugiere que el éxito académico refuerza las actitudes positivas hacia la educación, fortalece la percepción de control sobre los resultados escolares y se alinea con normas sociales que valoran la continuidad educativa. En contextos rurales, donde las presiones externas como la pobreza y el aislamiento geográfico a menudo debilitan la intención educativa, promover el éxito académico se posiciona como una estrategia clave para mitigar la deserción.

Igualmente, es relevante es el papel de la pobreza multidimensional, capturada por el Índice de Pobreza Multidimensional (IPM), que explicó hasta el 65% de la varianza en las tasas de aprobación ($R^2 = 0.6508$) en los modelos de árboles de decisión. Departamentos como Vichada y Guainía, con IPM de 70.2% y 49% respectivamente, presentan las tasas de deserción más altas (6.1% y 6.7% en 2023). Este resultado se alinea con la Teoría del Conflicto Social de Marx y Engels (1848), que plantea la educación como un terreno de lucha donde las desigualdades estructurales perpetúan la exclusión. En estas regiones, la pobreza —manifestada en viviendas inadecuadas, falta de acceso a agua potable y servicios de salud limitados— crea barreras sistémicas que afectan desproporcionadamente a los estudiantes rurales, perpetuando un ciclo de marginalización. Los datos sugieren que la deserción no es simplemente una decisión individual, sino un síntoma de inequidades socioeconómicas arraigadas que requieren intervenciones estructurales.

La elaboración de mapas de riesgo resalta las disparidades regionales, identificando a Guainía, Caquetá y Vichada como focos de alto riesgo de deserción. Estos hallazgos resuenan con la Teoría del Capital Social de Putnam (2000), que destaca la importancia de las redes comunitarias para mejorar los resultados educativos. En estas zonas de alto riesgo, el capital social debilitado —evidenciado por infraestructura deficiente, recursos comunitarios limitados y aislamiento— parece agravar las tendencias de abandono escolar. Por ejemplo, Vaupés y Amazonas, con tasas de matrícula de apenas 39.3% y 57.8%, reflejan cómo la dispersión geográfica y la falta de apoyo comunitario dificultan el acceso a la educación. En contraste, regiones como Nariño y Boyacá, con tasas de matrícula más altas (hasta 95.1% en Quindío) y mayores índices de aprobación, sugieren que un capital social más robusto actúa como un factor protector contra la deserción.

Al comparar estos resultados con la literatura existente, nuestro estudio coincide con Albor-Chadid et al. (2019), quienes señalaron que los factores socioeconómicos y educativos son impulsores principales de la deserción en Colombia. Sin embargo, nuestro uso de técnicas de aprendizaje automático, como árboles de decisión y modelos de regresión, añade una dimensión predictiva que no estaba presente en su trabajo, permitiendo identificar perfiles de riesgo específicos y patrones geográficos. De manera similar, Aravena Domich y Pérez Campillo (2022) destacaron el impacto del desempeño académico en la retención, pero el presente análisis amplía esta observación al cuantificar la fuerza de esta relación y vincularla con marcos teóricos. A diferencia de Morales Vega (2021), que se centró en la deserción relacionada con la pandemia, el actual enfoque abarca problemas estructurales persistentes, sugiriendo que, aunque las crisis intensifican la deserción, sus causas subyacentes están profundamente arraigadas en los contextos rurales.

Un resultado inesperado fue el bajo poder predictivo de las variables relacionadas con la violencia —desaparición forzada, reclutamiento forzado y violencia sexual—, con coeficientes de correlación cercanos a cero (por ejemplo, 0.028 para desaparición forzada y -0.005 para reclutamiento forzado). Dado el impacto histórico del conflicto armado en la educación colombiana (Palacios Tovar, 2017), la Teoría del Conflicto Social podría sugerir una correlación directa entre la violencia y la deserción al interrumpir el acceso y la seguridad. Sin embargo, estos modelos solo muestran una asociación leve, como la correlación de 0.094 entre reclutamiento forzado y repitencia. Esto podría indicar que la violencia afecta la deserción de manera indirecta, posiblemente a través de la inestabilidad económica o el desplazamiento, efectos que nuestros datos transversales no capturaron completamente. Estudios futuros podrían

explorar estas vías mediadoras utilizando datos longitudinales para alinear mejor los hallazgos empíricos con las expectativas teóricas.

Las implicaciones para las políticas públicas son claras. La fuerte relación entre la aprobación y la retención exige intervenciones educativas focalizadas, como capacitación docente y programas de tutoría, para impulsar el éxito académico en las escuelas rurales. Esto se alinea con la Teoría del Comportamiento Planificado, donde fortalecer el control percibido a través de sistemas de apoyo podría cambiar las intenciones de los estudiantes hacia la permanencia escolar. Al mismo tiempo, la relevancia predictiva del IPM subraya la necesidad de estrategias intersectoriales —mejorando el acceso a agua, nutrición y vivienda— para aliviar la carga de la pobreza en la educación. Estas medidas abordarían las inequidades estructurales destacadas por la Teoría del Conflicto Social, ofreciendo un enfoque integral para prevenir la deserción.

Las variaciones regionales orientan aún más estas recomendaciones. En departamentos de alto riesgo como Caquetá, donde factores relacionados con el conflicto, como el reclutamiento forzado (1287 casos), se cruzan con la pobreza, las intervenciones podrían combinar apoyo psicosocial con mejoras en la infraestructura. Por el contrario, regiones como Nariño, con tasas de deserción más bajas (1.9%), podrían servir como modelos para escalar programas comunitarios que aprovechen el capital social, como sugiere la teoría de Putnam. Este enfoque diferenciado garantiza que las políticas reflejen las realidades locales, aumentando su efectividad en el diverso paisaje rural de Colombia.

Las limitaciones del análisis merecen atención. El ajuste perfecto del modelo de regresión múltiple para el IPM y la deserción ($R^2 = 1.0000$, $MSE = 0.0000$) señala un posible sobreajuste, preocupante dado lo complejo de los fenómenos sociales. Esto sugiere que, aunque la pobreza es

un factor dominante, nuestro modelo puede exagerar su poder explicativo, requiriendo validación con datos no utilizados. Además, la dependencia de datos transversales limita la capacidad para rastrear la causalidad a lo largo del tiempo, una brecha que estudios longitudinales podrían cerrar. Estas restricciones matizan los hallazgos acá presentados, pero también destacan caminos para refinar modelos predictivos en investigaciones futuras.

Desde el punto de vista teórico, este estudio enriquece el discurso sobre la deserción al integrar múltiples perspectivas. La Teoría del Comportamiento Planificado enmarca la aprobación escolar como un motor conductual, la Teoría del Capital Social contextualiza las influencias comunitarias y la Teoría del Conflicto Social sitúa la deserción dentro de inequidades sistémicas. Juntas, ofrecen un marco integral que conecta la agencia individual con las fuerzas estructurales, avanzando más allá de los análisis de un solo factor predominantes en trabajos previos (por ejemplo, Utria, 2023). En términos prácticos, los mapas de riesgo y perfiles de estudiantes proporcionan herramientas accionables para educadores y tomadores de decisiones, alineándose con el objetivo del proyecto de informar intervenciones basadas en evidencia.

En síntesis, esta investigación revela que la deserción escolar en las zonas rurales de Colombia es el resultado de desafíos educativos y socioeconómicos entrelazados, con el éxito académico y la pobreza como predictores fundamentales. Aunque el impacto directo de la violencia parece limitado, sus efectos indirectos justifican una mayor exploración. Al fundamentar estos hallazgos en marcos teóricos robustos y datos empíricos, se contribuye a una comprensión más profunda de las dinámicas de la deserción, allanando el camino para estrategias focalizadas que promuevan la equidad educativa y la resiliencia en las comunidades rurales de Colombia.

Conclusiones

La presente investigación se propuso analizar los factores predictivos de la deserción escolar en zonas rurales de Colombia mediante un enfoque cuantitativo que integró análisis descriptivo, correlacional y predictivo. Utilizando datos oficiales del Ministerio de Educación Nacional (MEN), el Departamento Administrativo Nacional de Estadística (DANE) y otras fuentes gubernamentales, el estudio buscó identificar variables clave, examinar sus interacciones y desarrollar modelos predictivos para anticipar casos de riesgo. Los resultados ofrecen una visión multidimensional del fenómeno, destacando la interacción entre factores educativos, socioeconómicos y contextuales, y proporcionan una base empírica para orientar políticas educativas focalizadas. Esta sección presenta una síntesis de los hallazgos, sus implicaciones y su relación con los objetivos planteados, manteniendo claridad y precisión en la exposición.

Principales hallazgos.

El análisis identificó la aprobación escolar como el predictor más robusto de la permanencia educativa, con una correlación negativa de -0.92 respecto a la reprobación. Este resultado empírico, derivado de modelos estadísticos, indica que los estudiantes con mejor rendimiento académico tienen una menor probabilidad de abandonar la escuela, un hallazgo alineado con estudios previos como el de Gómez Triana (2016). Otro factor clave fue el Índice de Pobreza Multidimensional (IPM), que explicó hasta el 65% de la varianza en las tasas de aprobación según modelos de árboles de decisión ($R^2 = 0.6508$). Departamentos como Vichada (IPM 70.2%) y Guainía (IPM 49%) mostraron tasas de deserción elevadas (6.1% y 6.7% en 2023), evidenciando el peso de las condiciones socioeconómicas en la continuidad educativa.

Sorprendentemente, las variables de violencia (desaparición forzada, reclutamiento forzado y violencia sexual) no presentaron correlaciones significativas con la deserción en los

modelos predictivos, con coeficientes cercanos a cero (por ejemplo, 0.028 para desaparición forzada). Aunque se esperaba un impacto directo debido a la relevancia sociopolítica del conflicto armado, este resultado podría reflejar limitaciones en los datos, como subregistro o falta de desagregación temporal. Sin embargo, se detectó una asociación leve entre reclutamiento forzado y repitencia (0.094), sugiriendo efectos indirectos que ameritan mayor exploración. Por su parte, la tasa de matrícula mostró una correlación negativa moderada con la deserción (-0.31), indicando que un mayor acceso inicial reduce el abandono, especialmente en regiones como Nariño (matrícula 94.5%).

Interacciones entre factores y su interpretación.

Los datos revelaron interacciones no lineales entre los factores predictivos. Por ejemplo, en regiones con alta pobreza multidimensional y baja infraestructura educativa, como Vichada, el riesgo de deserción aumenta significativamente. El análisis de árboles de decisión identificó que dimensiones específicas del IPM, como el material inadecuado en viviendas ($R^2 = 0.1080$), tienen un impacto medible. Esta sinergia sugiere que las privaciones múltiples refuerzan las barreras a la permanencia escolar, una interpretación que va más allá de los resultados estadísticos y se basa en el contexto socioeconómico rural. Asimismo, la relación entre aprobación y repitencia (-0.71) apunta a la calidad educativa como un mediador crítico en el ciclo de abandono, un aspecto que requiere atención en políticas focalizadas.

Limitaciones metodológicas.

Los modelos predictivos enfrentaron desafíos que limitan su alcance. El modelo de regresión lineal múltiple para el IPM mostró un R^2 de 1.0000 y un MSE de 0.0000, lo que indica sobreajuste. Este resultado empírico sugiere que el modelo captura ruido específico de los datos de entrenamiento en lugar de generalizar patrones reales, un problema que podría mitigarse con

validación cruzada. Además, la falta de datos longitudinales impidió analizar la evolución temporal de la deserción, restringiendo la capacidad para establecer causalidad. La omisión de variables cualitativas, como la motivación estudiantil, también pudo haber afectado la precisión predictiva, destacando la necesidad de enfoques mixtos en futuros estudios.

Disparidades regionales.

Se constataron diferencias marcadas en las tasas de deserción entre regiones. Guainía (6.7%), Caquetá (6.6%) y Vichada (6.1%) registraron los niveles más altos, mientras que Nariño (1.9%) y Boyacá (2.1%) mostraron tasas bajas. Estos datos empíricos reflejan la influencia del contexto local: alta pobreza y baja matrícula en los primeros, frente a mejor infraestructura y cobertura en los segundos. En zonas afectadas por el conflicto, como Caquetá, la violencia podría mediar indirectamente el abandono a través de la inseguridad, aunque los modelos no lo capturaron directamente.

Implicaciones teóricas y prácticas.

Teóricamente, este estudio valida la relevancia de un enfoque multidimensional que integre factores educativos y socioeconómicos, extendiendo hallazgos de autores como Utria (2023). Prácticamente, los resultados sugieren priorizar la calidad educativa y abordar la pobreza multidimensional. Por ejemplo, la fuerte relación entre aprobación y permanencia respalda la implementación de programas pedagógicos, mientras que el impacto del IPM exige estrategias intersectoriales para mejorar condiciones básicas como agua y vivienda.

Contraste entre objetivos y resultados.

Tabla 8*Contraste entre Objetivos y Resultados*

Objetivo Específico	Grado de Cumplimiento	Justificación
Examinar factores predictivos	Cumplido	Se identificaron aprobación e IPM como predictores claves con evidencia estadística robusta.
Comparar patrones regionales	Cumplido	Se mapearon disparidades entre departamentos, respaldadas por datos de matrícula y deserción.
Construir modelos predictivos	Parcialmente cumplido	Los modelos mostraron capacidad variable; el sobreajuste limitó su utilidad práctica.
Formular recomendaciones estratégicas	Cumplido	Se propusieron acciones específicas, aunque podrían detallarse más por perfil de riesgo.

Reflexión ética.

El manejo de datos sensibles sobre violencia y pobreza plantea dilemas éticos. Aunque se utilizaron fuentes agregadas, la privacidad de las comunidades rurales debe protegerse para evitar estigmatización. Esto subraya la importancia de protocolos éticos rigurosos en investigaciones futuras, especialmente al tratar temas de desplazamiento o conflicto.

Recomendaciones de política pública.

Programas de Tutoría Personalizada: Implementar tutorías para estudiantes en riesgo de reprobación, en alianza con el ICBF, en regiones como Guainía.

Incentivos Económicos Condicionados: Coordinar con el DNP un programa de transferencias focalizado en familias pobres de Vichada, condicionado a la asistencia escolar.

Infraestructura Educativa: Expandir escuelas móviles en Vaupés, financiadas por el Fondo de Infraestructura Educativa.

Indicadores de Seguimiento: Establecer métricas como tasa de deserción anual por municipio para evaluar intervenciones.

Líneas para futuras investigaciones.

Análisis Longitudinal: Usar datos temporales para capturar dinámicas de deserción.

Variables Psicosociales: Incorporar motivación y apoyo familiar mediante métodos mixtos.

Modelos Avanzados: Aplicar redes neuronales para modelar relaciones no lineales.

En síntesis, este estudio aporta evidencia cuantitativa sobre los impulsores de la deserción escolar rural, destacando la aprobación y la pobreza como factores centrales. Aunque la violencia mostró un impacto limitado, su influencia indirecta merece mayor análisis. Estas conclusiones ofrecen lineamientos iniciales para políticas que promuevan la equidad educativa en Colombia, alineándose con el objetivo de mitigar un desafío estructural clave.

Recomendaciones

Los hallazgos de esta investigación ofrecen una base empírica sólida para orientar intervenciones que aborden la deserción escolar en las zonas rurales de Colombia. Conforme al cuarto objetivo específico del proyecto, las recomendaciones que se presentan a continuación se derivan directamente de la evidencia cuantitativa obtenida a través del análisis de datos y los modelos predictivos desarrollados. Estas propuestas están diseñadas para ser implementadas por actores clave como el Ministerio de Educación Nacional (MEN), las secretarías de educación departamentales y organizaciones no gubernamentales, y se enfocan en mitigar los factores de riesgo identificados, considerando las disparidades regionales y los perfiles de vulnerabilidad específicos. El análisis destacó la aprobación escolar, la pobreza multidimensional, la tasa de matrícula y las disparidades regionales como elementos críticos, los cuales sustentan las siguientes estrategias.

1. Mejorar la aprobación escolar mediante intervenciones pedagógicas focalizadas: El análisis reveló que la aprobación escolar es el predictor más robusto de la permanencia educativa, con una correlación negativa de -0.92 respecto a la reprobación y de -0.71 con la repitencia. Este hallazgo indica que los estudiantes con mejor rendimiento académico tienen una menor probabilidad de abandonar la escuela, lo que subraya la importancia de fortalecer la calidad educativa como estrategia preventiva. Para ello, se recomienda al MEN:

- Implementar programas de tutoría personalizada: Dirigidos a estudiantes en riesgo de reprobación o repitencia, especialmente en departamentos como Guainía, donde la repitencia alcanza el 19.4%. Estos programas podrían involucrar a docentes capacitados o voluntarios universitarios, enfocándose en asignaturas críticas como matemáticas y lenguaje, y operando en

horarios extracurriculares. La evidencia sugiere que un aumento en la aprobación reduciría significativamente las tasas de deserción, particularmente en regiones con perfiles de alto riesgo.

- **Capacitación docente en pedagogías contextualizadas:** Los resultados muestran que la aprobación explica una varianza sustancial en la permanencia escolar. Por ello, se propone formar a los docentes en estrategias pedagógicas adaptadas a las realidades rurales, como el uso de recursos locales y métodos interactivos. Esta medida podría implementarse en colaboración con el programa "Todos a Aprender", priorizando regiones con bajas tasas de aprobación, como Vichada (aprobación promedio del 70%).

- **Desarrollo de recursos educativos específicos:** Dado que la repitencia y la reprobación están inversamente relacionadas con la permanencia, se sugiere crear materiales didácticos que respondan a las necesidades culturales y lingüísticas de las comunidades rurales. Por ejemplo, en zonas como Vaupés, con una tasa de matrícula de solo 39.3%, estos recursos podrían incluir contenidos bilingües que faciliten el aprendizaje y reduzcan el riesgo de abandono.

- Estas intervenciones se alinean con los perfiles de riesgo identificados en los modelos predictivos, que señalan a los estudiantes con bajo rendimiento académico como los más vulnerables a la deserción.

2. **Reducir las barreras socioeconómicas asociadas a la pobreza multidimensional:** El Índice de Pobreza Multidimensional (IPM) emergió como un factor predictivo clave, explicando hasta el 65% de la varianza en las tasas de aprobación ($R^2 = 0.6508$) según los modelos de árboles de decisión. Regiones como Vichada (IPM 70.2%) y Guainía (IPM 49%) registran tasas de deserción de 6.1% y 6.7%, respectivamente, lo que evidencia cómo las condiciones

socioeconómicas adversas agravan el abandono escolar. Para mitigar este impacto, se propone una estrategia intersectorial con las siguientes acciones:

- Programas de transferencias condicionadas: Inspirados en modelos exitosos como los reportados por Fiszbein y Schady (2009), se recomienda al DNP y al MEN implementar incentivos económicos para familias en extrema pobreza, condicionados a la asistencia y permanencia escolar. En municipios como Taraira (Vaupés), donde la tasa de matrícula es de apenas 22.6%, este enfoque podría reducir la presión económica que lleva a los estudiantes a trabajar en lugar de estudiar, abordando directamente un perfil de riesgo caracterizado por alta pobreza y baja escolarización.

- Mejoramiento de infraestructura básica: El análisis identificó que dimensiones del IPM, como el material inadecuado en viviendas ($R^2 = 0.1080$), tienen una relación significativa con la deserción. Se sugiere priorizar inversiones en agua potable, saneamiento y vivienda en comunidades rurales de alto riesgo, coordinadas por el DNP y las alcaldías locales. Por ejemplo, en Vichada, donde el IPM alcanza el 70.2%, estas mejoras podrían aliviar las barreras logísticas y de salud que afectan la asistencia escolar.

- Fortalecimiento de programas de alimentación escolar: La correlación negativa entre matrícula y deserción (-0.31) sugiere que garantizar la nutrición es un incentivo para la permanencia. Se propone ampliar la cobertura del Programa de Alimentación Escolar (PAE) en zonas como Caquetá (deserción 6.6%), asegurando al menos una comida diaria para todos los estudiantes, lo que reduciría las tasas de abandono entre los perfiles más vulnerables identificados en los mapas de riesgo.

- Estas medidas abordan directamente las condiciones socioeconómicas que los modelos predictivos señalaron como determinantes clave de la deserción, ofreciendo soluciones diferenciadas para estudiantes en contextos de pobreza extrema.

3. Incrementar el acceso educativo para reducir la deserción inicial: La tasa de matrícula mostró una correlación negativa moderada con la deserción (-0.31), lo que indica que un mayor acceso inicial a la educación actúa como un factor protector. Departamentos como Vaupés (39.3%) y Amazonas (57.8%) presentan tasas de matrícula bajas, asociadas a mayores índices de abandono. Para revertir esta tendencia, se recomienda:

- Construcción de infraestructura educativa flexible: Los modelos predictivos identificaron la dispersión geográfica como una barrera crítica en regiones de alto riesgo. Se propone al MEN implementar escuelas móviles o aulas multigrado en zonas como Vaupés, donde la baja matrícula refleja dificultades de acceso. Estas soluciones podrían financiarse mediante el Fondo de Infraestructura Educativa, priorizando municipios con tasas de matrícula por debajo del 50%.

- Sistemas de transporte escolar gratuito: La evidencia cuantitativa señala que la distancia es un factor limitante en la permanencia escolar. Se sugiere establecer servicios de transporte en áreas rurales dispersas, como Pana Pana (Guainía), donde las largas caminatas desincentivan la asistencia. Esta medida podría coordinarse con las secretarías de educación departamentales, enfocándose en perfiles de riesgo caracterizados por aislamiento geográfico.

- Campañas de sensibilización comunitaria: Aunque no se cuantificó directamente, la baja matrícula en regiones como Guaviare (56.9%) sugiere una desvalorización de la educación. Se recomienda al MEN y a ONG lanzar campañas que promuevan la importancia de la escolarización, dirigidas a padres y líderes locales, utilizando medios como radio comunitaria

para llegar a zonas remotas. Esto fortalecería la matrícula inicial y reduciría la deserción en comunidades vulnerables.

Estas acciones buscan aumentar la cobertura educativa, un factor que los modelos predictivos identificaron como esencial para prevenir el abandono escolar en sus etapas tempranas.

4. Diseñar intervenciones diferenciadas según disparidades regionales: Los mapas de riesgo elaborados en el estudio evidenciaron disparidades regionales marcadas, con tasas de deserción más altas en Guainía (6.7%), Caquetá (6.6%) y Vichada (6.1%), frente a regiones como Nariño (1.9%) y Boyacá (2.1%). Estas diferencias reflejan contextos locales diversos que requieren enfoques personalizados. Por ello, se propone:

- Priorización de recursos en zonas de alto riesgo: Los modelos predictivos señalan a Guainía, Caquetá y Vichada como focos críticos. Se recomienda al MEN focalizar inversiones en estas regiones, destinando fondos adicionales para infraestructura, docentes y programas de apoyo. Por ejemplo, en Caquetá, con 1287 casos de reclutamiento forzado, se podrían combinar esfuerzos educativos con intervenciones psicosociales para atender este perfil específico de vulnerabilidad.
- Adaptación de estrategias al contexto local: En Vichada, donde la pobreza extrema predomina, las intervenciones deben integrar mejoras educativas con desarrollo socioeconómico, como las transferencias condicionadas ya mencionadas. En contraste, en Nariño, con una deserción baja, se podrían escalar programas comunitarios exitosos que aprovechen el capital social existente, como asociaciones de padres que promuevan la permanencia escolar.

- **Monitoreo y evaluación diferenciada:** Los datos cuantitativos sugieren que las tasas de deserción varían según las condiciones regionales. Se propone establecer indicadores específicos, como la tasa de deserción anual por municipio, para medir el impacto de las intervenciones y ajustarlas según los resultados. Esto podría implementarse mediante el Sistema de Información para el Monitoreo, la Prevención y el Análisis de la Deserción Escolar (SIMPAD), priorizando las regiones identificadas en los mapas de riesgo.

Este enfoque diferenciado asegura que las políticas públicas respondan a las particularidades de cada región, maximizando su efectividad en el diverso contexto rural colombiano.

Reflexiones Finales

Las recomendaciones aquí expuestas cumplen con el cuarto objetivo específico del proyecto al basarse exclusivamente en la evidencia cuantitativa derivada del análisis de datos y los modelos predictivos. La fuerte relación entre aprobación escolar y permanencia, el impacto del IPM, la influencia de la matrícula y las disparidades regionales ofrecen un marco claro para diseñar intervenciones estratégicas. Estas propuestas no solo abordan los factores de riesgo identificados, sino que también consideran los perfiles de vulnerabilidad específicos, desde estudiantes con bajo rendimiento en Guainía hasta comunidades aisladas en Vaupés.

La implementación de estas medidas requerirá una colaboración estrecha entre el MEN, el DNP, las secretarías de educación y la sociedad civil, así como un compromiso sostenido con la equidad educativa. Si se ejecutan de manera efectiva, estas estrategias tienen el potencial de reducir significativamente la deserción escolar en las zonas rurales, contribuyendo a romper los ciclos de pobreza y exclusión que afectan a las comunidades más vulnerables de Colombia. Además, los mapas de riesgo y los perfiles de estudiantes desarrollados en esta investigación

ofrecen herramientas prácticas para que los tomadores de decisiones prioricen recursos y adapten soluciones a las realidades locales, asegurando un impacto tangible en la educación rural.

Referencias Bibliográficas

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Albor-Chadid, L. I., Filut, D., Valero-Díaz, B., & Rodríguez-Burgos, K. (2019). Propuesta de modelo de innovación pedagógica para disminuir la deserción escolar en Colombia desde la experiencia de Israel. *Revista de Estudos de Literatura, Cultura e Alteridade-Igarapé*. <https://periodicos.unir.br/index.php/igarape/article/view/5420>
- Aravena Domich, M., & Pérez Campillo, C. C. (2022). La (no) deserción escolar. Análisis de estudiantes de 9.º grado en Colombia. *Revista Andina de Educación*, 5(1), 1–8. <https://doi.org/10.32719/26312816.2022.5.1.11>
- Bradley, M. (2020). Elevated school dropout rates in rural China. *Ballard Brief*. <https://ballardbrief.byu.edu/issue-briefs/elevated-school-dropout-rates-in-rural-china>
- Carvajal Valderrama, H. D. (2023). *Efectos de la suspensión de las aspersiones aéreas con glifosato sobre la deserción escolar en Colombia* [Tesis doctoral, Universidad de Los Andes]. <http://hdl.handle.net/1992/70190>
- Centro Nacional de Memoria Histórica. (s.f.). *Víctimas de desaparición forzada en el conflicto armado* [Conjunto de datos]. Observatorio de Memoria y Conflicto. <https://micrositios.centrodememoriahistorica.gov.co/observatorio/portal-de-datos/base-de-datos/>
- Centro Nacional de Memoria Histórica. (s.f.). *Víctimas de reclutamiento y utilización de niños, niñas y adolescentes* [Conjunto de datos]. Observatorio de Memoria y Conflicto. <https://micrositios.centrodememoriahistorica.gov.co/observatorio/portal-de-datos/base-de-datos/>

- Centro Nacional de Memoria Histórica. (s.f.). *Víctimas de violencia sexual en el conflicto armado* [Conjunto de datos]. Observatorio de Memoria y Conflicto.
<https://micrositios.centrodememoriahistorica.gov.co/observatorio/portal-de-datos/base-de-datos/>
- Cerón, A. C. S., Echavarría, G. H. V., & Noguera, C. A. R. (2024). La deserción escolar un problema generalizado en Colombia. *Ciencia y Educación*.
<https://doi.org/10.5281/zenodo.11407176>
- Congreso de la República de Colombia. (1994). *Ley 115 de 1994: Ley General de Educación*.
http://www.secretariasenado.gov.co/senado/basedoc/ley_0115_1994.html
- Congreso de la República de Colombia. (2001). *Ley 715 de 2001: Establece normas sobre el financiamiento de la educación*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=4452>
- Congreso de la República de Colombia. (2014). *Ley 1712 de 2014: Ley de Transparencia y Acceso a Información Pública*.
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882>
- Constitución Política de Colombia. (1991). Artículo 67, 68, 70.
http://www.secretariasenado.gov.co/senado/basedoc/constitucion_politica_1991_pr002.html
- Departamento Administrativo Nacional de Estadística (DANE). (2020). Inasistencia escolar en zonas rurales fue del 30% en 2020: DANE.
<https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/calidad-de-vida-ecv/encuesta-nacional-de-calidad-de-vida-ecv-2020>

Departamento Administrativo Nacional de Estadística (DANE). (2025). *Pobreza monetaria: Información pobreza monetaria nacional 2023*.

<https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-monetaria>

Departamento Administrativo Nacional de Estadística (DANE). (2025). *Índice de Pobreza Multidimensional – IPM-2024*. Microdatos DANE.

<https://microdatos.dane.gov.co/index.php/catalog/860/get-microdata>

Departamento Administrativo Nacional de Estadística (DANE). (2025). *Pobreza multidimensional*. <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-multidimensional>

Departamento Nacional de Planeación (DNP). (s.f.). *Terridata: Plataforma de información territorial*. <https://terridata.dnp.gov.co/index-app.html/descargas>

Departamento Nacional de Planeación (DNP). (2025). *TerriData: Descargas*.

<https://terridata.dnp.gov.co/index-app.html/descargas>

Diosa Guiral, H., Ospina López, J. E., Reina Morales, H. D., & López Bedoya, G. (2024).

Investigación sobre la deserción escolar en Colombia mediante el uso de técnicas de machine learning. *Repositorio Institucional UNIREMINGTON*.

<https://repositorio.uniremington.edu.co/handle/123456789/2130>

Elmasri, R., & Navathe, S. B. (2016). *Fundamentals of database systems* (7th ed.). Pearson.

<https://www.pearson.com/en-us/subject-catalog/p/fundamentals-of-database-systems/P200000003546/9780137502523>

Eslamian, A., Arshi, M., Tahmasebi, S., Alipour, F., Rafiey, H., & Karimijavan, G. (2023).

Understanding dropout stabilization and the factors affecting the return to primary

- school: A qualitative study. *Journal of Medicine and Life*, 16(7), 990–997.
<https://doi.org/10.25122/jml-2022-0022>
- Ferrari, A., & Russo, M. (2016). *Introducing Microsoft Power BI*. Microsoft Press.
https://download.microsoft.com/download/0/8/1/0816F8D1-D1A5-4F60-9AF5-BC91E18D6D64/Microsoft_Press_ebook_Introducing_Power_BI_PDF_mobile.pdf
- Gómez Molina, S., Valencia-Arías, A., Saldarriaga Ríos, J. G., Vélez Holguín, R. M., & Soto Giraldo, J. O. (2022). Deserción escolar de niños y niñas en Colombia en tiempos de pandemia. *Telos Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 24(3), 628–642. <https://dialnet.unirioja.es/servlet/articulo?codigo=8586707>
- Gómez Triana, B. (2016). *Deserción escolar en áreas rurales de Colombia*. Universidad Javeriana.
<https://repository.javeriana.edu.co/bitstream/handle/10554/21107/GomezTrianaBelenEugenia2016.pdf?sequence=3&isAllowed=y>
- González, F. (2016). Jóvenes vulnerables, educación rural y construcción de la paz en Colombia. *NOREF*. https://noref.no/publication-documents/building-the-peace-rural-education-and-conflict-in-colombia/Gonzalez_NOREF_Report_NRC_Educacion_Colombia_web.pdf
- Hernández, R., Fernández, C., & Baptista, P. (2018). *Metodología de la investigación* (6ª ed.).
<https://dialnet.unirioja.es/servlet/libro?codigo=775008>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer. <https://www.statlearning.com/>
- Kandel, S., Paepcke, A., Hellerstein, J. M., & Heer, J. (2011). Wrangler: Interactive visual specification of data transformation scripts. *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems, 3363–3372.

<https://dl.acm.org/doi/10.1145/1978942.1979444>

La República. (2023). En 2022, 23,7% de los jóvenes en zonas rurales no asistió a instituciones educativas. <https://www.larepublica.co/economia/caracteristicas-de-la-educacion-en-zonas-rurales-3719292>

Laboratorio de Economía de la Educación, Universidad Javeriana. (2024). Educación rural en crisis: Menos del 50% de estudiantes logra grado 11.

<https://cambiocolombia.com/educacion/crisis-educacion-rural-infraestructura-carencia-docentes-desercion-escolar>

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley.

<https://onlinelibrary.wiley.com/doi/book/10.1002/9781119482260>

Marx, K., & Engels, F. (1848). *Manifiesto del Partido Comunista*.

<https://www.marxists.org/espanol/m-e/1840s/48-manif.htm>

McKinney, W. (2010). Data structures for statistical computing in Python. En S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61).

SciPy. <https://proceedings.scipy.org/articles/Majora-92bf1922-00a.pdf>

Ministerio de Educación Nacional (MEN). (2025). *Matrícula en educación en preescolar, básica y media hasta 2023*. Datos Abiertos Colombia. https://www.datos.gov.co/Educacion/MEN_MATRICULA_EN_EDUCACION_EN_PREESCOLAR-B-SICA-Y-/sty3-c395/data_preview

Ministerio de Educación Nacional. (2022). *Deserción escolar en Colombia: Análisis, determinantes y política de acogida, bienestar y permanencia* [Nota técnica].

https://www.mineducacion.gov.co/1780/articles-363488_recurso_34.pdf

Ministerio de Educación Nacional. (s.f.). *Encuesta Nacional de Deserción Escolar*.

https://www.mineducacion.gov.co/1621/articles-293664_archivo_pdf_resultados_ETC.pdf

Morales Vega, J. F. (2021). *Deserción escolar en el marco de la Pandemia del COVID-19 en Colombia* [Tesis de maestría, Universidad de Los Andes].

<http://hdl.handle.net/1992/55077>

Naciones Unidas. (1948). *Declaración Universal de Derechos Humanos*. Artículo 26.

<https://www.hchr.org.co/publicaciones/declaracion-universal-de-derechos-humanos/>

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO).

(1989). *Convenio sobre los Derechos del Niño*. Artículos 28 y 29.

<https://www.un.org/es/events/childrenday/pdf/derechos.pdf>

Palacios Tovar, M. (2017). *Revisión de la literatura de la relación entre conflicto armado y educación en Colombia*. Universidad Militar Nueva Granada.

<https://repository.unimilitar.edu.co/server/api/core/bitstreams/79fd3895-4468-4f8c-a200-5dbe7c1a5654/content>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Research, 12, 2825–2830.

<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–

1227. <https://www.science.org/doi/10.1126/science.1213847>

Presidencia de la República de Colombia. (2007). *Decreto 2888 de 2007: Reglamenta la atención educativa en áreas dispersas*.

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=26017>

Presidencia de la República de Colombia. (2015). *Decreto 103 de 2015: Reglamenta el uso de datos abiertos*. <https://www.apccolombia.gov.co/normativa/decreto-no-103-de-2015>

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3–13.

https://www.researchgate.net/publication/220282831_Data_Cleaning_Problems_and_Current_Approaches

Rodríguez, M. F., Arenas, C. M., Valenzuela, J. P., & Nussbaum, M. (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), 10103–10130. <https://doi.org/10.1007/s10639-022-11515-5>

Sectorial. (2024). 473.780 menores abandonaron sus estudios en el 2023.

<https://sectorial.co/educacion-basica-y-preescolar/473-780-menores-abandonaron-sus-estudios-en-el-2023/>

SITEAL. (2022). *Sistema de Información de Tendencias Educativas en América Latina*.

<https://siteal.iiep.unesco.org/indicadores>

Sosa Chinome, J. A., & Vargas Rodríguez, G. R. (2022). Consecuencias de la deserción escolar de niños y jóvenes víctimas del conflicto armado en Colombia.

<https://dialnet.unirioja.es/servlet/articulo?codigo=9085630>

- UNICEF. (2024). Dejar de ir a la escuela porque se seca el río: Más de 420.000 niños se han visto afectados por la sequía en la Amazonia. <https://elpais.com/america-futura/2024-11-08/dejar-de-ir-a-la-escuela-porque-se-seca-el-rio-mas-de-420000-ninos-se-han-visto-afectados-por-la-sequia-en-la-amazonia.html>
- United Way Colombia. (2022). Deserción escolar: Desafío de la educación en Colombia. <https://unitedwaycolombia.org/desercion-escolar-desafio-de-la-educacion-en-colombia/>
- Utria, Y. (2023). Educación rural y desarrollo sostenible en Colombia: Perspectivas y desafíos. <https://doi.org/10.56219/rgp.vi47.2354>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace. <https://dl.acm.org/doi/book/10.5555/1593511>
- Velasco Quintero, S. J., & Castro Garzón, V. O. (2023). *La deserción en la educación rural en Colombia. Un problema multidimensional. Una revisión documental 2007—2020*. Universidad Pedagógica Nacional. <http://repository.pedagogica.edu.co/handle/20.500.12209/18627>
- Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1), 3–28. <https://doi.org/10.1198/jcgs.2009.07098>
- Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, 59(11), 56–65. <https://cacm.acm.org/research/apache-spark/>
- Zapata, J. (2018). La educación: Un grave problema de la ruralidad colombiana. <https://aneia.uniandes.edu.co/la-educacion-un-grave-problema-de-la-ruralidad-colombiana/>

Apéndices

Apéndice A

Visual del Código Python con Análisis de Datos

<https://luchaconm31.github.io/github.io-TrabajoGrado/>

Apéndice B

Tablero Power BI para Análisis de Datos

<https://app.powerbi.com/view?r=eyJrIjoiOTBmZDUwNWUtYjNiNi00OWY3LTlmYTMtZDU2YWM4ODFhY2NiIiwidCI6ImZjMDA1NDdhLTI0YmItNGU0Zi05ZDYxLTczZmNhNWViOWRmMyIsImMiOiR9>