









## IA aplicada a la identificación de características de ocupación de menores en hogares colombianos para detectar posible trabajo infantil

### AI Applied to the Identification of the Occupational Characteristics of Minors in Colombian Households to Detect Potential Child Labor

### IA aplicada a identificação de características de ocupação de menores em lares colombianos para detectar possível trabalho infantil

Roxana-María Romero-Luna<sup>1</sup>    
Hugo-Armando Ordoñez-Erazo<sup>2</sup>    
Carlos-Alberto Cobos-Lozada<sup>3</sup>  

**Recibido:** 2 de julio de 2024

**Aceptado:** 30 de agosto de 2024

**Para citar este artículo:** Romero-Luna, R. M., Ordoñez-Erazo, H. A., y Cobos-Lozada, C. A. (2024). IA aplicada a la identificación de características de ocupación de menores en hogares colombianos para detectar posible trabajo infantil. *Revista Científica*, 50(2), 55-70. <https://doi.org/10.14483/23448350.22623>

#### Resumen

El trabajo infantil es una antigua problemática mundial que tiende a ser normalizada en ciertos contextos culturales, sociales y económicos. Colombia, un país en vías de desarrollo que ha sido afectado por el conflicto armado y su complicada topografía, enfrenta también la falta de oportunidades laborales para madres o jefes de hogar. Estos factores, junto con los desafíos en el acceso a la educación, hacen que el país sea vulnerable al trabajo infantil. En este artículo se utilizan datos de la Encuesta de Calidad de Vida 2022 del Departamento Administrativo Nacional de Estadística (DANE), específicamente del módulo de trabajo infantil. Con estos datos, se propone un enfoque de inteligencia artificial (IA) basado en algoritmos de aprendizaje automático para clasificar la ocupación de menores de edad en los hogares colombianos. Se encontró que el modelo óptimo se obtiene al aplicar el algoritmo de clasificación DecisionTreeClassifier en datos debidamente procesados y utilizando validación cruzada estratificada. Al analizar los atributos seleccionados, que son los más importantes y mejor rankeados, se observa que la ocupación de los menores está fuertemente influenciada por las características económicas y la composición del hogar. Se concluye que las técnicas de IA son cruciales para identificar los factores que inciden en problemas sociales como el trabajo infantil, y que pueden servir de apoyo para los entes gubernamentales en la implementación de estrategias de mitigación y prevención.

**Palabras clave:** aprendizaje automático, inteligencia artificial, trabajo infantil

1. MSc Universidad del Cauca (Popayán-Cauca, Colombia). [romeroluna@unicauca.edu.co](mailto:romeroluna@unicauca.edu.co).
2. PhD Universidad del Cauca (Popayán-Cauca, Colombia). [hugoordonez@unicauca.edu.co](mailto:hugoordonez@unicauca.edu.co).
3. PhD Universidad del Cauca (Popayán-Cauca, Colombia). [ccobos@unicauca.edu.co](mailto:ccobos@unicauca.edu.co).

## Abstract

Child labor is an age-old global issue that tends to be normalized in certain cultural, social, and economic contexts. Colombia, a developing country affected by armed conflict and its challenging topography, also faces a lack of employment opportunities for mothers or heads of households. These factors, along with the challenges in accessing education, make the country vulnerable to child labor. This article uses data from the 2022 Quality of Life Survey of the National Administrative Department of Statistics (DANE), specifically from the child labor module. With these data, an artificial intelligence (AI) approach based on machine learning algorithms is proposed to classify the occupation of minors in Colombian households. It was found that the optimal model is obtained by applying the DecisionTreeClassifier algorithm to duly processed data and using stratified cross-validation. By analyzing the selected attributes, which are the most important and highest ranked, it is observed that the occupation of minors is strongly influenced by household economic characteristics and composition. It is concluded that AI techniques are crucial for identifying the factors that influence social problems such as child labor, and that they can serve as support for government entities in implementing mitigation and prevention strategies.

**Keywords:** artificial intelligence, child labor, machine learning

## Resumo

O trabalho infantil é uma antiga problemática mundial que tende a ser normalizada em certos contextos culturais, sociais e econômicos. A Colômbia, um país em desenvolvimento que foi afetado pelo conflito armado e sua complicada topografia, também enfrenta a falta de oportunidades de emprego para mães ou chefes de família. Esses fatores, juntamente com os desafios no acesso à educação, tornam o país vulnerável ao trabalho infantil. Neste artigo, utilizam-se dados da Pesquisa de Qualidade de Vida de 2022 do Departamento Administrativo Nacional de Estatística (DANE), especificamente do módulo de trabalho infantil. Com esses dados, propõe-se uma abordagem de inteligência artificial (IA) baseada em algoritmos de aprendizado de máquina para classificar a ocupação de menores de idade nos lares colombianos. Foi constatado que o modelo ideal é obtido ao aplicar o algoritmo de classificação DecisionTreeClassifier em dados devidamente processados e utilizando validação cruzada estratificada. Ao analisar os atributos selecionados, que são os mais importantes e melhor classificados, observa-se que a ocupação dos menores é fortemente influenciada pelas características econômicas e pela composição do lar. Conclui-se que as técnicas de IA são cruciais para identificar os fatores que incidem em problemas sociais como o trabalho infantil e que podem apoiar os órgãos governamentais na implementação de estratégias de mitigação e prevenção.

**Palavras-chaves:** inteligência artificial, aprendizagem de máquina, trabalho infantil

---

## INTRODUCCIÓN

Según la [UNICEF \(2006\)](#), se entiende por *niño* todo ser humano menor de 18 años de edad, salvo que, en virtud de la ley que le sea aplicable, haya alcanzado antes la mayoría de edad. El *trabajo infantil* se define como cualquier actividad laboral que prive a los niños de su infancia, potencial y dignidad y sea perjudicial para su desarrollo físico y mental ([Organización Internacional del Trabajo, 2024a](#)). Estas actividades pueden ser peligrosas y dañinas para los niños en términos mentales, físicos, sociales o morales, además de interferir con su educación al impedirles asistir a la escuela, forzarlos a abandonarla prematuramente, o exigirles que combinen la asistencia escolar con un trabajo excesivamente largo y pesado.

El trabajo infantil es un problema global y arraigado que a menudo se normaliza en ciertos contextos culturales, sociales y económicos ([Lucas & Mendoza, 2024](#)). Por ejemplo, a finales de la década de 1860

([Bermeosolo, 1962](#)), la venta masiva de periódicos fue un fenómeno relativamente común en varios países. En sus inicios y durante décadas, esta actividad fue realizada principalmente por niños. El trabajo infantil sigue siendo un problema complejo a nivel mundial, afectando a 160 millones de niños, de los cuales 63 millones son niñas y 97 millones son niños (*i.e.*, casi uno de cada diez niños). De estos, aproximadamente 79 millones están involucrados en trabajos peligrosos ([UNICEF, 2021](#)).

En América Latina, las familias tienen dificultades para adquirir alimentos, ropa y otros bienes y servicios. Como resultado, los miembros de la familia optan por trabajar desde una edad temprana para contribuir con recursos al hogar ([Acuña Ramírez et al., 2021](#)), combinando el estudio con el trabajo en el mejor de los casos, o abandonando definitivamente el primero.

Los últimos datos disponibles (2020) indican que 8.2 millones de niños entre los 5 y los 17 años trabajan en América Latina y el Caribe. Una gran parte de estos menores son adolescentes masculinos, mientras que el 33 % corresponde a niñas. El trabajo infantil está presente en áreas tanto rurales como urbanas, y un 48.7 % de los niños labora en el sector agrícola. Un poco menos de la mitad de quienes se ven involucrados en el trabajo infantil lo hacen dentro del ámbito familiar. Más del 50 % de los menores realiza labores peligrosas que afectan su salud, educación y bienestar ([Organización Internacional del Trabajo, 2024b](#)).

Colombia es un país en desarrollo y marcado por el conflicto armado, lo que ha causado el desplazamiento de población. Además, su variada topografía limita la movilidad de los habitantes, dificultando el acceso de los menores a la educación. Asimismo, el país alberga diversos grupos étnicos, donde el trabajo infantil es una parte de la realidad cotidiana debido a sus creencias y tradiciones ([Raza & Riaz, 2023](#)). Tampoco se puede olvidar el impacto de la migración de venezolanos ([Franco Ordóñez & Vera, 2022](#); [Organización Integral para las Migraciones, 2024](#); [Zewdie et al., 2024](#)). Teniendo en cuenta estos factores, podemos decir que Colombia es un país propenso al trabajo infantil. Esto es evidente, pues es muy común encontrar niños trabajando en los semáforos, dedicándose a la venta de todo tipo de alimentos o artículos y la limpieza de parabrisas, entre otras actividades.

En Colombia, el trabajo infantil constituye una violación a los derechos fundamentales de los niños, niñas y adolescentes, pues dificulta su desarrollo integral y equilibrado, y no les permite crecer en un entorno familiar y comunitario de bienestar, afecto y comprensión, tal como lo estipula la Ley 1098 de 2006 ([Instituto Colombiano de Bienestar Familiar, 2006](#)).

El objetivo general de las políticas públicas en Colombia es garantizar que todos los niños, niñas y adolescentes crezcan en entornos familiares protectores que los alejen del trabajo infantil, y que los adolescentes trabajadores puedan ejercer sus derechos dentro de un marco de trabajo protegido ([Gobierno de Colombia, 2018](#); [Ministerio del Trabajo, 2017](#)).

Aunque las investigaciones que relacionan la inteligencia artificial (IA) con temas sociales como el trabajo infantil son recientes, estudios como el de [Chocobar \(2021\)](#) en Argentina han utilizado modelos predictivos para entender y predecir este fenómeno. Por ejemplo, [Guevara et al. \(2024\)](#) aplicaron una técnica de IA neutrosófica para evaluar la prevalencia y las consecuencias del trabajo infantil, encontrando, en efecto, un alto índice de prevalencia y diversas consecuencias negativas para la salud y el desarrollo de los niños. Aunque existen algunos trabajos que utilizan técnicas de IA para abordar el trabajo infantil, los hallazgos son escasos.

En este artículo se propone la aplicación de la IA mediante algoritmos de aprendizaje automático para clasificar la ocupación de menores de edad en los hogares colombianos, utilizando datos de la Encuesta de Calidad de Vida del DANE ([Departamento Administrativo Nacional de Estadística, 2022c](#)). El objetivo es identificar las características socioeconómicas que contribuyen a la incorporación temprana de menores

al trabajo y brindar apoyo a las entidades gubernamentales en la formulación de políticas públicas. Estas políticas, como las mencionadas por [Ministerio del Trabajo \(2017\)](#), el [Gobierno de Colombia \(2018\)](#) y [Pacto Global Red Colombia \(2024\)](#), buscan contribuir a la prevención y erradicación del trabajo infantil y sus peores formas en las empresas y sus cadenas de suministro y áreas de influencia, además de proteger a los adolescentes trabajadores bajo regímenes de protección especial. Estas iniciativas son lideradas por el Ministerio de Trabajo y el Pacto Global Red Colombia. El objetivo es diseñar estrategias que permitan prevenir y erradicar progresivamente el trabajo infantil. Por lo tanto, es importante proveer estas herramientas como una alternativa para contribuir a la formulación de políticas públicas efectivas en este ámbito.

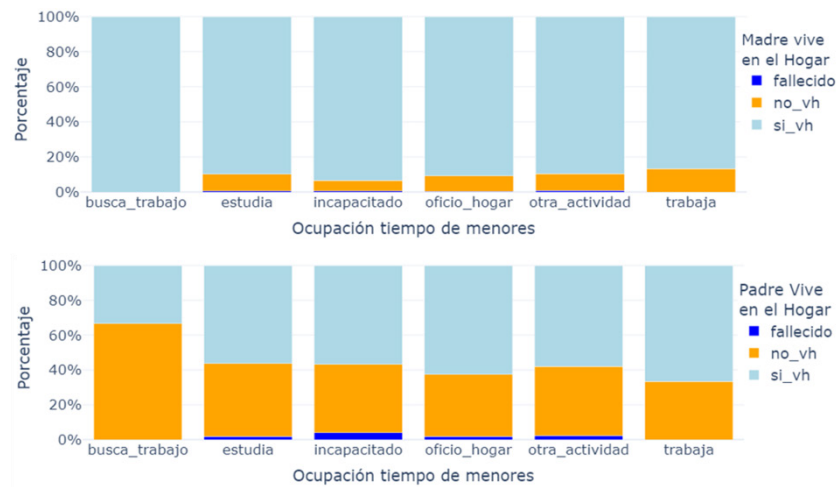
Además, este trabajo busca responder al interés de entidades internacionales como la Organización Internacional del Trabajo (OIT) y el Fondo Internacional de Emergencia de las Naciones Unidas para la Infancia (UNICEF) ([Cuello et al., 2024](#); [Organización Internacional del Trabajo, 2016](#)), así como a los esfuerzos del gobierno colombiano a través del Sistema de Información Integrado para el Registro y la Erradicación del Trabajo Infantil y sus Peores Formas (SIRITI) ([Colombia Potencia de la Vida, 2024](#)), los Comités de Erradicación del Trabajo Infantil (CIETI) ([Rubio, 2024](#)) y el ICBF ([Instituto Colombiano de Bienestar Familiar, 2020](#)), con su Proceso Administrativo de Restablecimiento de Derechos (PARD) ([Organización Internacional del Trabajo, 2016](#)) para mitigar el trabajo infantil. Las técnicas de IA son valiosas para analizar e identificar los factores que influyen en la ocupación de menores en los hogares colombianos, considerando sus contextos sociales y económicos. Los modelos de aprendizaje automático pueden procesar grandes volúmenes de datos eficientemente, mejorando la precisión de las predicciones y las clasificaciones. Estos modelos pueden actualizarse con nuevos datos sin necesidad de reprogramación manual.

Este documento está organizado de la siguiente manera. Se inicia con un escenario de motivación, se expone la metodología implementada y luego se presentan los pasos de la misma, que incluye el entendimiento del problema; la adquisición, la comprensión y la preparación de los datos; y el modelamiento con los respectivos resultados.

## MOTIVACIÓN Y TRABAJOS RELACIONADOS

Tanto la familia como la cultura son fundamentales en la formación de los patrones psicosociales de una persona, pues influyen en su comportamiento durante toda su vida ([Canales y Alva, 2023](#)). La [Figura 1](#) muestra cómo la situación de la madre o el padre en el hogar inciden en la ocupación de los menores de edad.

De los menores que buscan trabajo, el 100 % tiene madre y el 67 % no vive con su padre, lo que sugiere que una gran parte de ellos proviene de hogares con madres cabeza de hogar. Los niños con madres cabeza de familia y sin padre en el hogar enfrentan numerosas desventajas, como limitaciones económicas que afectan su acceso a necesidades básicas y oportunidades educativas. Además, las madres solteras suelen experimentar altos niveles de estrés y carga emocional, lo que impacta su bienestar y el de sus hijos ([Williams, 2021](#)). Sin embargo, de los menores que trabajan, el 87 % tiene madre y el 67 % padre. Esto sugiere que los padres influyen en su ingreso al mercado laboral, reflejando patrones culturales y sociales. La familia, como primer entorno social, influye en la percepción del trabajo como una necesidad, ya sea por supervivencia o exigencia socioeconómica ([Narváez, 2019](#)). El estudio de Maldonado Montaña y Tobón Sarmiento (2021) en Aratoca, Santander, muestra que los padres justifican el trabajo infantil con la falta de ingresos y lo promueven para inculcar disciplina y alejar a los jóvenes de actividades inapropiadas o ilícitas.

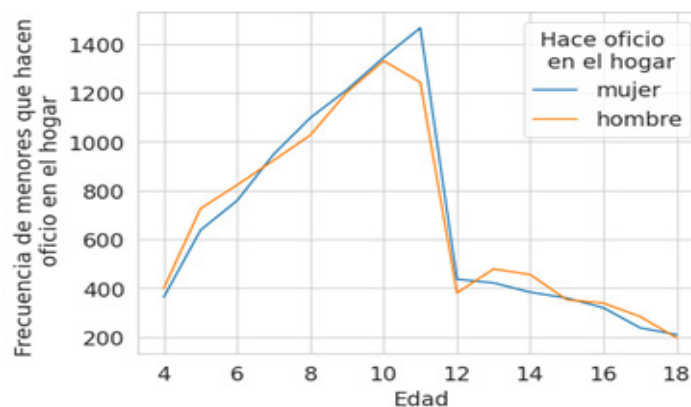


**Figura 1.** Porcentaje de ocupación a tiempo parcial de menores según la situación del padre y la madre en el hogar, i.e., fallecido, no vive en el hogar (no\_vh) o sí vive en el hogar (sí\_vh)

En cuanto a los menores dedicados a los oficios del hogar, se hace evidente que, en general, esta población cuenta con padre y madre. En Colombia, es común que las niñas y adolescentes se dediquen a la mayor parte de las actividades domésticas; se les asigna el cuidar a los hermanos menores y encargarse de otras tareas de la casa.

En la [Figura 2](#) se puede apreciar cómo la población de mujeres menores es la que más realiza trabajos domésticos. En particular, esto se hace evidente en el rango de edad de 10 a 12 años. La mayoría de los menores que realizan labores en sus casas oscila entre los 5 y los 11 años.

Algunas de las tareas del hogar incluyen limpiar la vajilla; poner la mesa; limpiar, secar, alisar y remendar la ropa; asear la vivienda; hacer las camas; limpiar y fregar el piso; y vaciar los desechos. Estas actividades domésticas pueden interferir con la educación y el desarrollo personal de los niños, y constituyen una forma de trabajo infantil no remunerado que afecta a un porcentaje considerable de niños y adolescentes en el país ([Departamento Administrativo Nacional de Estadística, 2024a](#)).

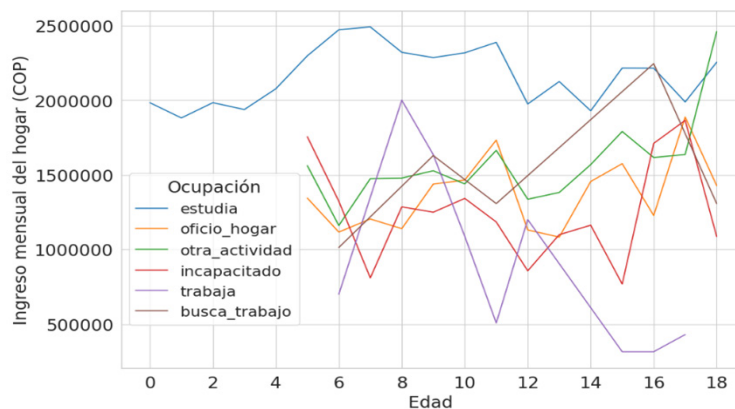


**Figura 2.** Distribución de la frecuencia de menores que realizan oficios en el hogar según su edad y sexo

Según el boletín técnico del módulo de Trabajo Infantil de la Gran Encuesta Integrada de Hogares elaborado por el DANE para el trimestre octubre-diciembre de 2022, a nivel nacional había 369 mil niños y adolescentes de 5 a 17 años que estaban trabajando. De este grupo, en áreas rurales y centros poblados se encontraba un 55.6 %, lo equivalente a 206 mil menores, mientras que el 44.4 % residía en centros urbanos municipales, sumando 164 mil personas. En términos de género, durante el período de referencia, el 69.5 % de los trabajadores menores de edad eran hombres (257 mil personas) y el 30.5 % eran mujeres (113 mil personas). En cuanto a la distribución por edad, el mayor porcentaje de trabajo infantil se observó en el grupo de 15 a 17 años, con un 70.1 %, mientras que el grupo de 5 a 14 años constituyó el 29.9 % ([Departamento Administrativo Nacional de estadística, 2022a](#)).

En Colombia, la falta de ingresos en los hogares, causada por la carencia de oportunidades y el limitado nivel académico de los adultos responsables del hogar, es un factor que contribuye significativamente al problema del trabajo infantil. A este respecto, mediante el método histórico-hermenéutico y con un enfoque cualitativo e inductivo, [García \(2021\)](#) describió y estudió cómo el trabajo infantil impacta el desarrollo de los objetivos sostenibles en Barranquilla, concluyendo que el empleo infantil en el área urbana central de la ciudad está influenciado por la falta de oportunidades educativas, la escasez de ingresos económicos y la urgencia de apoyar la estabilidad económica del hogar.

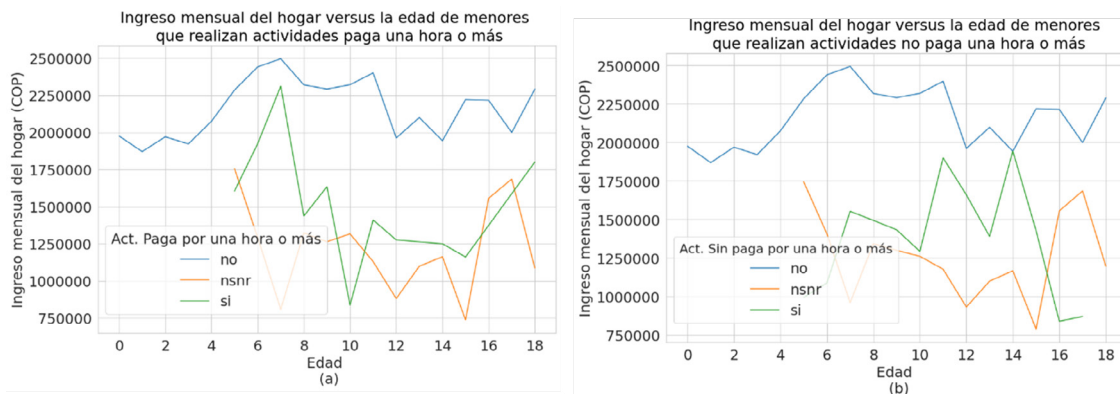
La [Figura 3](#) muestra que la ocupación de los menores entre los 0 y los 14 años es exclusivamente académica si el ingreso mensual del hogar es mayor o igual a 2 millones de pesos, mientras que, si el ingreso es menor, los infantes o adolescentes, la mayoría entre los 6 y los 14 años, realizan oficios en el hogar, trabajan, buscan trabajo o realizan otras actividades.



**Figura 3.** Relación entre la edad de los menores, su ocupación y el ingreso mensual del hogar en pesos colombianos

Asimismo, en la [Figura 4](#) se evidencia que los menores entre 6 y 16 años realizan actividades con o sin paga por una o más horas cuando el ingreso mensual del hogar está por debajo de los dos millones de pesos.

Entre los ejemplos posibles de esa actividad paga están la prestación de servicios a domicilio, la preparación de alimentos para la venta y la venta de artesanías, entre otros. Entre las actividades no pagas, según el manual de conceptos básicos de la ECV 2022, está la del trabajador familiar sin remuneración, que no recibe salario en dinero ni en especie después de trabajar por lo menos por una hora a la semana en una empresa económica de propiedad de una persona emparentada con él/ella y que reside en el mismo hogar ([Departamento Administrativo Nacional de Estadística, 2022b](#)).



**Figura 4.** Relación entre la edad de los menores, la actividad que realizan con o sin paga por una hora o más y el ingreso mensual del hogar en pesos colombianos

Las repercusiones del trabajo infantil incluyen, entre otras, el aumento del desempleo, limitaciones relacionadas con los ingresos económicos, la dificultad para generar ahorros, la continuidad del ciclo de pobreza, el deterioro del capital humano y el incremento de la deserción escolar. Por ello, es crucial aunar esfuerzos hacia la eliminación del trabajo infantil.

Entre las estrategias desarrolladas por la OIT están el Programa Internacional para la Erradicación del Trabajo Infantil (IPEC) y la Iniciativa Regional América Latina y el Caribe Libre de Trabajo Infantil. Esta última es una alianza liderada por 30 países de la región y que cuenta con la participación activa de organizaciones de empleadores y de trabajadores. Juntos materializan el compromiso de prevenir y erradicar el trabajo infantil en la región a través de políticas, estrategias y acciones innovadoras y urgentes ([Organización Internacional del Trabajo, 2016](#)).

Para cumplir con este objetivo, se creó la CIETI, una instancia encargada de la prevención y la erradicación del trabajo infantil y la protección del adolescente trabajador a nivel municipal o departamental, cuyas funciones son implementar políticas públicas; asesorar, coordinar y proponer programas y estrategias para prevenir y erradicar el trabajo infantil; fortalecer la coordinación y la concertación entre las instituciones presentes; y elaborar un plan de acción para prevenir y erradicar el trabajo infantil y proteger al adolescente trabajador ([Ministerio del Trabajo, 2024](#)).

La literatura que aborda el problema del trabajo infantil es extensa y, por lo general, intenta identificar tanto las causas principales como los factores que influyen en este fenómeno. [Macías \(2021\)](#) examina el impacto de las inundaciones en el trabajo infantil y la educación en hogares rurales de Colombia, explorando si la inclusión financiera mitiga estos efectos. Con otro enfoque, el artículo de [Rubio \(2024\)](#) analiza la implementación de políticas públicas para erradicar el trabajo infantil en el Caquetá, utilizando una metodología cualitativa centrada en la articulación de entidades y la protección de los derechos de los niños, mientras que [Franco Ordóñez y Vera \(2022\)](#) estudian el trabajo infantil y la migración en la frontera entre el departamento de Norte de Santander y Táchira (Venezuela), utilizando métodos cuantitativos para proteger los derechos de los niños.

En este contexto, este artículo, que se centra en el uso de técnicas de IA, puede ser útil para la toma de decisiones orientadas a la prevención del trabajo infantil. Esto se debe a que analiza datos socioeconómicos, información sobre los servicios del hogar y las características y la composición de los hogares en áreas y grupos vulnerables, lo que facilita la identificación de riesgos. El modelo presentado

permite predecir incidencias al identificar patrones históricos y anticipar dónde es más probable que ocurra el trabajo infantil. También permite el monitoreo y la evaluación del apoyo a políticas y programas implementados por el Estado, ajustándolos en tiempo real. Por último, optimiza los recursos, dirigiéndolos a las áreas y grupos con mayor riesgo, maximizando así el impacto de las intervenciones estatales y de las organizaciones internacionales, y aprovechando el potencial de la IA para complementar los trabajos citados anteriormente.

## METODOLOGÍA

En este artículo se empleó la metodología CRISP-DM (Proceso Estándar de la Industria para la Minería de Datos). Este método proporciona una estructura organizada para desarrollar proyectos de minería y ciencia de datos, donde se hace uso de técnicas y algoritmos de IA y aprendizaje automático (ML) para analizar y extraer información útil de grandes conjuntos de datos, guiando a los equipos a través de varias etapas clave ([Sharma et al., 2017](#)). La metodología consta de seis fases: la comprensión del problema o negocio, la adquisición y la comprensión de datos, la preparación de datos, el modelado, la evaluación del modelo y el despliegue.

### Comprensión del problema

En la fase de entendimiento del problema se revisaron investigaciones y la literatura sobre el trabajo infantil de fuentes nacionales e internacionales. Se analizaron estadísticas globales y locales con el objetivo de comprender la magnitud del problema, identificar las causas y revisar las políticas públicas en Colombia. Entre los factores que más inciden en la incorporación temprana de los menores al mercado laboral, se destaca el económico. Según el DANE, en el último trimestre del 2022, los motivos principales del trabajo infantil fueron el deseo de tener ingresos propios (37.1 %) y la necesidad de contribuir a la economía familiar (33.3 %). Estudios locales como los de [García \(2021\)](#) y [Maldonado Montaña y Tobón Sarmiento \(2021\)](#) confirman que la falta de ingresos es un factor recurrente en la justificación del trabajo infantil. Además [Chocobar \(2021\)](#) evaluó modelos predictivos de este fenómeno y determinó que variables como la edad, el sexo y la región de residencia también son influyentes. Es así como, desde nuestra estrategia basada en algoritmos de ML, buscamos analizar y predecir las características que inciden en la ocupación de los menores en hogares colombianos, en aras de establecer relaciones con los estudios revisados.

### Fuentes de datos

La adquisición y comprensión de datos comenzó con el acceso a los datos de la ECV 2022, proporcionados por el DANE a través de su catálogo de datos en la página del Archivo Nacional de Datos (ANDA) ([Departamento Administrativo Nacional de Estadística, 2022c](#)). Se utilizaron los módulos de Trabajo Infantil, Características y Composición del Hogar, y Servicios del Hogar, de los cuales se obtuvieron inicialmente 138 708 registros y 200 atributos en formato CSV. Se seleccionaron los atributos que tuviesen, como máximo, el 5 % de datos nulos, y se eliminaron los atributos directorio y secuencia\_ecuesta, dado que no representaban información relevante. Como resultando, se obtuvo un archivo con 138 708 registros y 36 atributos. En la [Tabla 1](#) se presentan estos atributos y su descripción.



**Tabla 1.** Atributos del dataset y su descripción

Variable	Descripción
No_Cuartos_Hog	Esto abarca la sala, el comedor y otras habitaciones como la biblioteca, el salón de estar y el cuarto de invitados.
niños_comparten_cuart_adul	Algunos/as de los niños, niñas o adolescentes del hogar entre 0 y 17 años comparten el cuarto con un adulto.
pagan_energ	Si el hogar paga por el servicio de electricidad que consume.
serv_gas_publ	Si el hogar tiene conexión a una red de gas natural por ducto.
tipo_serv_sanitario	El medio por el cual se eliminan las excretas.
pagan_serv_alcant	Si el hogar paga por el servicio de alcantarillado.
como_eli_basu	El manejo de los desechos del hogar.
clasif_basu	Prácticas de clasificación que realiza el hogar respecto a la eliminación.
reutilizan_agua	Hogares que realizan acciones de ahorro de agua mediante la reutilización del agua.
recolectan_aguaLL	Hogares que realizan acciones ahorradoras de agua mediante la recolección de aguas lluvias.
pagan_serv_acued	Si el hogar paga por servicio de acueducto.
obt_agua	Procedencia del agua que el hogar utiliza para preparar los alimentos que requieren cocción.
ubic_sumin_agua	Localización el suministro del agua para el hogar.
tratam_agua_cosu	Si el hogar somete el agua que le suministran a algún sistema de purificación o potabilización.
lugar_preparan_alim	Si el hogar dispone de un espacio en la vivienda destinado exclusivamente a la preparación de alimentos.
tipo_ener_comb_usan_cocinar	El tipo de energía o combustible usado para cocinar.
tienen_serv_tele	Si el hogar cuenta con servicio telefónico, fijo o celular.
hg_realiza_act_produc	Si el hogar cuenta con un predio para el desarrollo de una actividad económica.
i_hogar	Ingreso mensual total del hogar.
i_ugasto	Ingreso mensual total de la unidad de gasto.
Percápita	Ingreso per cápita (ingreso mensual por persona en el hogar).
ingr_mens_uni_gast	Ingreso mensual total de la otra unidad.
cant_personas_hogar	Número de personas que componen el hogar.
ocupa_mptiempo	En qué actividad ocupó el encuestado la mayor parte del tiempo: estudiando, trabajando, en oficios del hogar, buscando trabajo, incapacitado o en otro tipo de actividades.
realiza_act_paga_uh_mash	Realizó actividad por una hora o más con pago.
traba_negocio_rec pago	Trabajo realizado en algún negocio familiar por el cual se recibe pago.
trabau_uh_mh_sinpag	Realizó actividad por una hora o más sin pago.
oficio_hogar	Aquellos que durante la semana de referencia se dedicaron principalmente a la gestión y el cuidado de su propio hogar.
sexo	Sexo del encuestado al nacer.
edad	Edad de la persona encuestada.
paren_jefeH	Parentesco con el jefe del hogar.
Lugar_Naci	Lugar de nacimiento de la persona encuestada.
padre_vive_h	Si el padre vive en el hogar.
madre_vive_h	Si la madre vive en el hogar.
se_considera	Caracterización étnica de la población.

Fue necesario imputar datos en los atributos relacionados en la [Tabla 2](#) con su respectivo porcentaje de datos nulos.

**Tabla 2.** Atributos con datos nulos

Atributo	Datos nulos
ubic_sumin_agua	4 %
tipo_energ_comb_usar_cocina	0.3 %
realiza_act_paga_uh_mas	0.51 %
trabaja_negocio_recpago	0.6 %

El procedimiento de imputación de los datos se dividió como se muestra a continuación.

Los valores faltantes en los atributos *ubic\_sumin\_agua* y *tipo\_energ\_comb\_usar\_cocina* fueron imputados utilizando un modelo de clasificación basado en el algoritmo DecisionTreeClassifier. Para ello, se entrenó el modelo con los datos completos disponibles, que incluían ejemplos sin valores nulos para estos atributos. Una vez entrenado, el modelo predijo los valores faltantes con base en las relaciones aprendidas entre los atributos. Así, los valores imputados se ajustaron de manera coherente al resto de los datos. La evaluación de este modelo se presenta en la [Tabla 3](#).

**Tabla 3.** Evaluación de DecisionTreeClassifier para imputación de datos nulos

Atributo	Exactitud	Matriz de Confusión
ubic_sumin_agua	0.82	<pre>0.8247896521411529 [[ 21522  1822   67]  [  4146  9086  315]  [   227   399 2231]]</pre>
tipo_energ_comb_usar_cocina	0.97	<pre>0.9755867837094526 [[ 424   7   0  15   0  11   0   0]  [  2 9954   0   2   0   0   0   0]  [  0   0  63   1   0   3   0   0]  [ 22   6   2 8444   1 216  23   0]  [  0   0   0   4  75   3   1   0]  [ 11   0   0 225   2 6673  29   0]  [  0   0   0  18   0  44 259   0]  [  0   0   0   0   0   0   0  3]]</pre>

La *exactitud* representa el porcentaje de predicciones correctas. Esta métrica es buena en ambos casos, siendo especialmente alta para el atributo *tipo\_energ\_comb\_usar\_cocina*, con un valor de 0.97, como se evidencia en su respectiva matriz de confusión. Para la imputación de los datos en el atributo *ubic\_sumin\_agua*, la exactitud también es buena (0.82).

Para el caso de los atributos *realiza\_act\_paga\_uh\_mas*, *trabaja\_negocio\_recpago* y *trabaj\_uh\_mh\_sinpaga*, se remplazaron los datos nulos por 'no sabe no responde' (nsnr), haciendo un mapeo mediante una función en código Python, la cual, con el comando `fillna(3)` remplazaba los datos nulos por el valor entero 3, que está asociado a la etiqueta nsnr.

## Preparación de datos

Los datos categóricos fueron codificados con el OneHotEncoder de scikit-learn, obteniendo 138 708 registros y 114 atributos. Además, se estandarizaron los datos con MinMaxScaler, de la misma librería. Se hizo la respectiva revisión de valores atípicos y se determinó la distribución de las clases en la variable objetivo *ocupa\_mptiempo*, la cual se muestra en la [Figura 5](#).

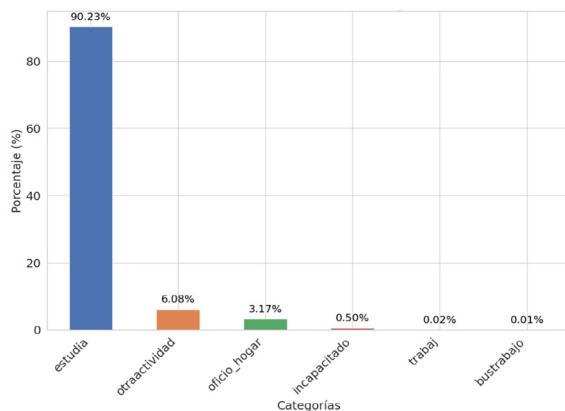


Figura 5. Distribución de clases en la variable objetivo

## Modelado

El modelo de IA presentado en este artículo se basa en algoritmos de ML que buscan ajustar tanto métodos sencillos de clasificación (e.g., regresión logística y el DecisionTreeClassifier) como modelos de ensamblaje (RandomForest, XGBoost y AdaBoost) a los datos de la ECV después del procesamiento. Este modelo se compone de tres partes: datos y procesamiento, algoritmos de ML y evaluación mediante validación cruzada estratificada. La [Figura 6](#) ilustra el modelo propuesto. La validación cruzada estratificada se realizó dividiendo el conjunto de datos en 10 pliegues ( $n\_splits=10$ ). Para ello, se utilizó la estrategia StratifiedKFold y la función *cross\_val\_score* de la biblioteca scikit-learn, evaluando los modelos con diversas métricas que se presentarán más adelante. A diferencia de la validación cruzada estándar, StratifiedKFold garantiza que cada pliegue mantenga una proporción de clases similar a la del conjunto de datos original. Esto es importante para evitar que algunas clases estén poco representadas o falten en ciertos pliegues a causa del desbalance de los datos.

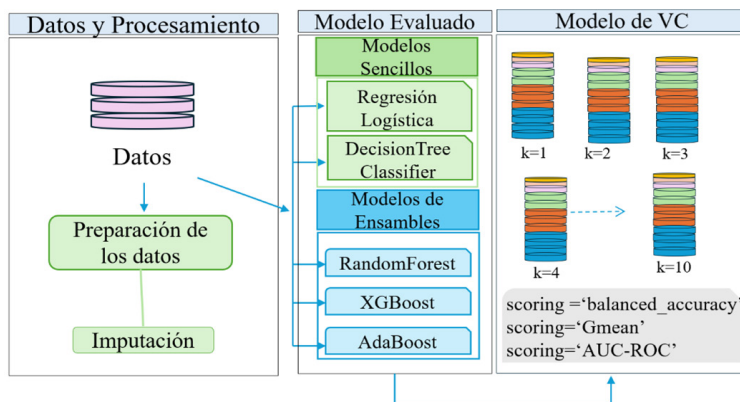


Figura 6. Componentes del modelo

Para evaluar el rendimiento del modelo propuesto, se utilizaron las métricas presentadas en la [Tabla 4](#).

**Tabla 4.** Métricas de evaluación

Métrica	Descripción
Exactitud Equilibrada ( <i>balanced_accuracy_score</i> )	Se calcula como el promedio de las tasas de aciertos para cada clase en el conjunto de datos (Brownlee, 2020).
AUC-ROC ( <i>roc_auc_score</i> )	El AUCROC indica la probabilidad de que un clasificador puntúe una instancia positiva con un valor más alto que una negativa elegida al azar (Brownlee, 2020).
Media geométrica ( <i>geometric_mean_score</i> )	Esta medida busca maximizar y equilibrar la precisión en cada una de las clases. En problemas multiclase, se calcula como la raíz superior del producto de la sensibilidad de cada clase (Barandela <i>et al.</i> , 2003).

## RESULTADOS

Como se mostró en la [Figura 6](#), la distribución de las clases en la variable objetivo está desbalanceada. En este contexto, las métricas clave para evaluar el rendimiento del modelo son la exactitud equilibrada (*balanced\_accuracy\_score*), AUC-ROC y G-mean (Barandela *et al.*, 2003).

En la [Tabla 5](#) se muestra el rendimiento de los modelos cuando se evalúan con validación cruzada estratificada para las métricas *balanced\_acuracy\_score*, *roc\_auc* y *Gmean*.

**Tabla 5.** Evaluación de la validación cruzada

Algoritmo	Validación cruzada scoring = <i>balanced_accuracy_score</i> (Promedio)	Validación cruzada scoring = <i>roc_auc</i> (Promedio)	Validación cruzada scoring = <i>Gmean</i> (Promedio)
DecisionTreeClassifier	<b>0.8002</b>	<b>0.8750</b>	<b>0.8644</b>
Random Forest	0.7009	0.9867	0.8031
XGBoost	0.7380	0.9742	0.8127
AdaBoost	0.3286	0.8797	0.5256
Regresión logística	0.4724	0.8893	0.6317

Considerando los resultados presentados en la [Tabla 5](#), y dado que las métricas que priorizan las clases minoritarias son de particular interés en este contexto, el modelo con mejor rendimiento es el DecisionTreeClassifier, con una exactitud equilibrada de 0.8002 y un G-Mean de 0.8031. Además, este modelo ofrece la solución más sencilla debido a su interpretabilidad; es considerado una caja de cristal y generaliza mejor el problema en comparación con los otros algoritmos, que al ser ensambles de árboles se vuelven más complejos para su interpretación. Esto es coherente con el principio de la navaja de Ockham (Witten *et al.*, 2017). La lista ordenada de importancia de los primeros 30 atributos para este modelo se presenta en la [Figura 7](#).

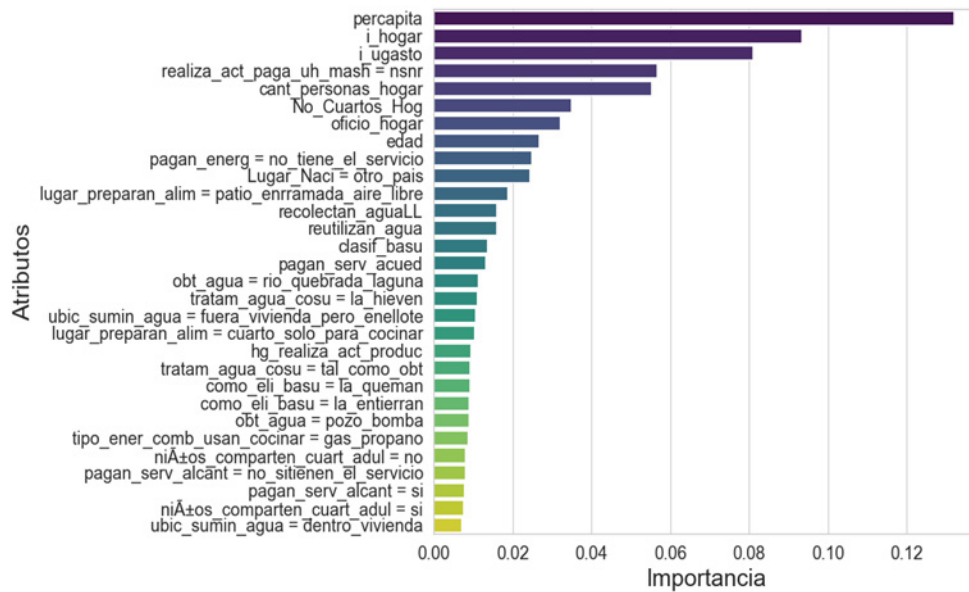


Figura 7. Lista de los primeros 30 atributos del modelo DecisionTreeClassifier en orden de importancia

En la Figura 8 se muestra la puntuación promedio de la métrica AUC-ROC para cada pliegue, destacando el valor máximo (0.91), alcanzado en el pliegue 7.

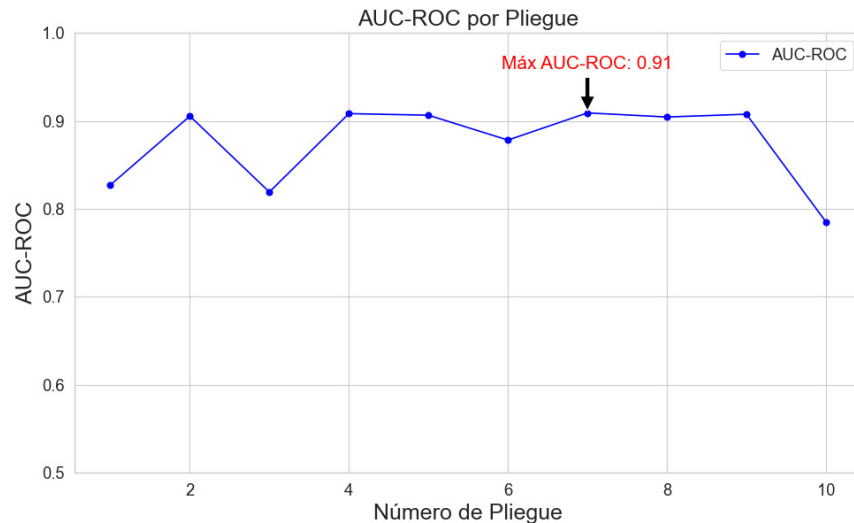


Figura 8. Puntuación promedio de la métrica AUC-ROC para cada pliegue en el modelo DecisionTreeClassifier

## CONCLUSIONES

Al implementar la estrategia propuesta, se encontró que el modelo óptimo es aquel que se obtiene al ajustar los datos debidamente procesados al algoritmo DecisionTreeClassifier, usando validación cruzada estratificada con  $n\_splits=10$ ,  $shuffle=True$  y  $random\_state=42$ . Este modelo alcanzó una exactitud

equilibrada de 0.8002, un AUC-ROC de 0.8750 y un G-mean de 0.8644. La puntuación promedio máxima de 0.91 para AUC-ROC se obtuvo en el pliegue 7. La estrategia permitió verificar que la ocupación de los menores en un hogar está fuertemente influenciada por las características económicas y la composición del mismo.

Los atributos seleccionados resaltan la importancia de la presencia parental en el hogar. Como señala la [UNICEF \(2022\)](#), la ausencia de uno o ambos padres va en detrimento de los beneficios de una paternidad activa, como lo son como una mayor autoestima, mejores habilidades sociales, buen desempeño escolar y el bienestar psicológico. Además, los atributos relacionados con las condiciones económicas de los hogares reflejan una relación con la ocupación de menores. Los atributos del acceso al agua y las condiciones de saneamiento también son cruciales, pues la calidad de vida impacta la salud y el desarrollo. La [Organización de las Naciones Unidas \(2024\)](#) subraya que el manejo eficiente del agua y el saneamiento son cruciales para disminuir enfermedades, así como para mejorar la salud y la productividad.

Las técnicas de IA son clave para identificar los factores que afectan problemas sociales como el trabajo infantil. Estos modelos, al entrenarse con grandes conjuntos de datos, mejoran la precisión de las predicciones y se actualizan automáticamente con nueva información, facilitando el análisis de datos complejos. Este artículo demostró cómo, tras procesar 138 000 registros y 114 atributos, se identificaron características relevantes para clasificar la ocupación de menores en los hogares colombianos.

Para mejorar el rendimiento del modelo e identificar con mayor precisión la importancia de las características por clase, se pueden utilizar técnicas de balanceo como SMOTE y métodos explicativos como SHAP. SMOTE ayuda a corregir el desbalanceo de clases mediante la generación de ejemplos sintéticos, mientras que SHAP desglosa la relevancia de las características por clase, facilitando la comprensión de su influencia en las predicciones del modelo. Esto destaca, de nuevo, cómo la IA ofrece una diversidad de técnicas y alternativas para abordar problemas sociales tan complejos como el objeto de este análisis. Este artículo demuestra cómo el ML puede proporcionar información valiosa para organizaciones y entidades gubernamentales que busquen reducir las estadísticas de participación de menores en trabajos que van en contra de su bienestar, sus derechos y su dignidad.

## AGRADECIMIENTOS

Al doctorado en Ciencias de la Computación de la Facultad de Ingeniería de la Universidad del Cauca.

## REFERENCIAS

- Acuña Ramírez, J. A., Gálvez Briones, P. E., Obando Peralta, E. C. (2021). Gasto público en educación y su impacto en el trabajo infantil en Cajamarca y Huancavelica. *Quipukamayoc*, 29(61), 47-56. <https://doi.org/10.15381/quipu.v29i61.20530>
- Barandela, R., Sánchez, J. S., García, V., Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3), 849-851. [https://doi.org/10.1016/S0031-3203\(02\)00257-1](https://doi.org/10.1016/S0031-3203(02)00257-1) [Get rights and content](#)
- Bermeosolo, F. (1962). La opinión pública norteamericana y la guerra de los Estados Unidos contra España. *Revista de Estudios Políticos*, 123, 219-234.
- Brownlee, J. (2020). *Imbalanced classification with Python. Machine Learning Mastery* (2da ed.). Guiding Tech Media.
- Canales, F. F. R., Alva, A. D. P. U. (2023). Lo social, histórico cultural y psicosocial: triada confusa, ambigua e imprecisa. *Revista Latinoamericana Ogmios*, 3(6), 101-112. <https://doi.org/10.53595/rlo.v3.i6.057>

- Colombia Potencia de la Vida (2024). *Sistema de información integrado para el registro y la erradicación del trabajo infantil y sus peores formas*. <https://app2.mintrabajo.gov.co/siriti/index.aspx#>
- Chocobar, E. B. (2021). *Prediciendo trabajo infantil: comparación de técnicas de econometría tradicional y machine learning* [Tesis de maestría, Universidad de San Andrés]. <https://dspaceapi.live.udesa.edu.ar/server/api/core/bitstreams/89f6b638-b3a4-42df-b6ef-a4fe99d577c1/content>
- Cuello, Y. I. L., Vidal, J. E., Brito, K. Y. S. (2024). Situación actual del trabajo infantil en Latinoamérica: una mirada desde los constructos educativos. *Encuentros. Revista de Ciencias Humanas, Teoría Social y Pensamiento Crítico*, 21, 104-115. <https://doi.org/10.5281/zenodo.11118092>
- Departamento Administrativo Nacional de Estadística (2022a). *Boletín técnico gran encuesta integrada de hogares (GEIH) módulo de trabajo infantil (MTI). Octubre – Diciembre 2022*. [https://www.dane.gov.co/files/investigaciones/boletines/ech/trabajo-infantil/bol\\_trab\\_inf\\_oct2022-dic2022.pdf](https://www.dane.gov.co/files/investigaciones/boletines/ech/trabajo-infantil/bol_trab_inf_oct2022-dic2022.pdf)
- Departamento Administrativo Nacional de Estadística (2022b). *Encuesta Nacional de Calidad de Vida – ECV*. <https://microdatos.dane.gov.co/index.php/catalog/793/related-materials>
- Departamento Administrativo Nacional de Estadística (2022c). *Encuesta Nacional de Calidad de Vida – ECV 2022*. Retrieved from [https://microdatos.dane.gov.co/index.php/catalog/793/data-dictionary/F21?file\\_name=Servicios%20del%20hogar](https://microdatos.dane.gov.co/index.php/catalog/793/data-dictionary/F21?file_name=Servicios%20del%20hogar)
- Departamento Administrativo Nacional de Estadística (2024). *Boletín técnico trabajo infantil (TI). Octubre – Diciembre 2023*. <https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIHTI-oct2022-dic2023.pdf>
- Franco Ordóñez, J. L., Vera, X. F. (2022). *Análisis del fenómeno migratorio: "trabajo infantil en Colombia"* [Tesis de especialización, Corporación Universitaria Minuto de Dios]. <https://repository.uniminuto.edu/server/api/core/bitstreams/5ead36aa-bf0e-4e63-9115-0e3bd5704036/content>
- García, M. R. (2021). Trabajo infantil frente al desarrollo de los objetivos sostenibles en la ciudad de Barranquilla localidad Cevillar. *Erg@ Omnes*, 13(1), 221-234.
- Gobierno de Colombia (2018). *Política nacional de infancia y adolescencia*. [https://www.icbf.gov.co/sites/default/files/politica\\_nacional\\_de\\_infancia\\_y\\_adolescencia\\_2018\\_-\\_2030.pdf](https://www.icbf.gov.co/sites/default/files/politica_nacional_de_infancia_y_adolescencia_2018_-_2030.pdf)
- Guevara, J. A. U., Cevallos, D. P. G., Flores, M. J. C., Cisneros, E. N. P. (2024). Método neutrosófico multicriterio para determinar prevalencia del trabajo infantil y la vulneración de los derechos fundamentales de los niños, niñas y adolescentes. *Revista Asociación Latinoamericana de Ciencias Neutrosóficas*, 31, 23-34.
- Instituto Colombiano de Bienestar Familiar (2006). *Código de la infancia y la adolescencia*. <https://www.icbf.gov.co/sites/default/files/codigoinfancialey1098.pdf>
- Instituto Colombiano de Bienestar Familiar (2020). *Trabajo infantil en Colombia*. [https://www.icbf.gov.co/system/files/infografiatrabajo infantil\\_vf.pdf](https://www.icbf.gov.co/system/files/infografiatrabajo infantil_vf.pdf)
- Lucas, J. B., Mendoza, E. H. (2024). Patrones culturales que produce la familia en relación al trabajo infantil. 593 *Digital Publisher CEIT*, 9(3), 892-905.
- Macías, C. B. (2021). *Child labour, rainfall shocks, and financial inclusion: Evidence from rural Colombia* [Tesis de maestría, Universidad del Rosario]. <https://repository.urosario.edu.co/server/api/core/bitstreams/742aaf89-1084-4fd5-986c-0447a2482180/content>
- MaldonadoMontaño, E., TobónSarmiento, R. E. (2021). *Patrones culturales relacionados con el trabajo infantil y adolescente, residentes en el Municipio de Aratoca Santander* [Tesis de maestría, Universidad Nacional Abierta y a Distancia]. <https://repository.unad.edu.co/bitstream/handle/10596/40546/Emaldonadom.pdf?sequence=3&isAllowed=y>
- Ministerio del Trabajo (2024). *Sistema de información integrado para el registro y la erradicación del trabajo infantil y sus peores formas*. <https://app2.mintrabajo.gov.co/siriti/index.aspx#>

- Ministerio del Trabajo (2017). *Línea de política pública para la prevención y erradicación del trabajo infantil y la protección integral al adolescente trabajador*. [https://www.mintrabajo.gov.co/documents/20147/58804143/Politica+publica+pdf\\_version+final\\_02042018.pdf](https://www.mintrabajo.gov.co/documents/20147/58804143/Politica+publica+pdf_version+final_02042018.pdf)
- Narváez, D. (2019). Aplicación de la ética en la familia y el hogar, El entorno familiar. *Aplicación de La Ética Familiar*, 1 (10), a15.
- Organización Integral para las Migraciones (2024). *Los migrantes y refugiados venezolanos en Colombia generan un impacto económico equivalente a 529,1 millones de dólares*. <https://www.iom.int/es/news/estudio-de-la-oim-los-migrantes-y-refugiados-venezolanos-en-colombia-generan-un-impacto-economico-equivalente-5291-millones-de-dolares>
- Organización Internacional del Trabajo (2016). *Iniciativa regional América Latina y el Caribe Libre de Trabajo Infantil*. <https://www.ilo.org/es/projects-and-partnerships/projects/iniciativa-regional-america-latina-y-el-caribe-libre-de-trabajo-infantil>
- Organización Internacional del Trabajo (2024a). Trabajo Infantil. <https://www.ilo.org/topics/child-labour/what-child-labour>
- Organización Internacional del Trabajo (2024b). *Trabajo infantil en América Latina y el Caribe*. Retrieved from <https://www.ilo.org/es/migration-stub-4877/trabajo-infantil-en-america-latina-y-el-caribe>
- Organización de las Naciones Unidas (2024). *Desafíos globales. Agua*. <https://www.un.org/es/global-issues/water>
- Pacto Global Red Colombiana (2024). *Red colombiana contra el trabajo infantil*. <https://www.pactoglobal-colombia.org/trabajo-infantil/red-colombia-contra-el-trabajo-infantil.html>
- Raza, F., Riaz, N. (2023). Deconstructing the socialization of female child in filter's child labor in America. *Global Sociological Review*, VIII(I), 427-432. [https://doi.org/10.31703/gsr.2023\(VIII-I\).42](https://doi.org/10.31703/gsr.2023(VIII-I).42)
- Rubio, J. M. M. (2024). Trabajo infantil: un reto institucional en el Caquetá. *Nova et Vetera*, 33, 1-18. <https://doi.org/10.22431/25005103.919>
- Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., Martin, L. (2017). *An agile group aware process beyond CRISP-DM: A hospital data mining case study* [Artículo de conferencia]. ICCDA '17: International Conference on Compute and Data Analysis. <https://doi.org/10.1145/3093241.3093273>
- UNICEF (2006). *Convención sobre los derechos del niño*. <https://www.un.org/es/events/childrenday/pdf/derechos.pdf>
- UNICEF (2021). *Trabajo infantil: estimaciones mundiales para 2020, tendencias y camino a seguir*. <https://www.ilo.org/publications/major-publications/child-labour-global-estimates-2020-trends-and-road-forward>
- UNICEF (2022). *Los beneficios de una paternidad activa para niños y niñas*. <https://www.unicef.org/chile/historias/los-beneficios-de-una-paternidad-activa-para-ni%C3%B1os-y-ni%C3%B1as>
- Williams, L. (2021). *Los efectos en los niños nacidos de madres solteras*. [https://www.ehowenespanol.com/efectos-ninos-nacidos-madres-solteras-info\\_374862/](https://www.ehowenespanol.com/efectos-ninos-nacidos-madres-solteras-info_374862/)
- Witten, I. H., Frank, E., Hall, M. A. (2017). *Data mining* (4a ed.). Morgan Kaufmann.
- Zewdie, B. M., Berlie, A. B., & Mersha, L. A. (2024). Causes of child labor and working conditions in Ethiopia: Evidence from temporary inter-rural child labor migrants from Sekela district. *Comparative Migration Studies*, 12(1), e28.

