# Predicción de la pobreza de los hogares en Colombia para el año 2018 usando aprendizaje automático[[1]](#footnote-1)

Palabras claves: *Análisis de pobreza, línea de pobreza, medición de pobreza,*

*encuesta de hogares y datos de hogares*

Códigos *JEL: I32, C81*

## Introducción

### Contexto

La pobreza es uno de los mayores desafíos de política social que enfrentan los países a nivel mundial, según datos del World Bank Group (2024) a nivel mundial el 8.5% (o 700 millones de personas) viven con menos de 2.5 dólares por persona al día; para América Latina y el Caribe la cifra de pobreza y pobreza extrema para el año 2023 fue de 10,6% y 27,3% respectivamente, lo que se traduce en 66 millones de personas (en condición de pobreza monetaria) y 172 millones de personas (en condición de pobreza monetaria extrema) (CEPAL, 2024). Para Colombia las cifras de pobreza monetaria y monetaria extrema para el año 2023 fueron de 33,0% y 11,4% respectivamente (DANE, 2024)

### Antecedentes

La literatura teoría y empírica sobre los estudios de pobreza monetaria y multidimensional abundan en cientos de artículos escrito por cientos de académicos experto en la temática (identificación, medición y evaluación de la pobreza sea directa o indirecta), se destacan autores como (Alkire & Foster, 2011; Banerjee & Duflo, 2012; Ravallion, 1994; Ray, 2002; Sen, 2000).

El economista Amartya Sen (2000) una de las mayores eminencias en temas de medición del bienestar y la pobreza, identifica dos formas de pobreza una indirecta que hace referencia a la pobreza monetaria[[2]](#footnote-2) o por ingresos, esta esencialmente se basa en una línea de pobreza, la CEPAL (2018) dice que “ La línea de pobreza representa un valor monetario en que se consideran dos componentes: el costo de adquirir una canasta básica de alimentos y el costo de los demás bienes y servicios, expresado sobre la base de la relación entre el gasto total y el gasto en alimentos”, la otra forma es la directa que se refiere a la pobreza multidimensional[[3]](#footnote-3) y esta es más amplia y tiene en cuenta muchas más dimensiones que el ingreso o consumo de los hogares, como por ejemplo la estructura del hogar en términos de condiciones de vida (habitabilidad, acceso a servicios esenciales como el agua, o alcantarillado, etc.) (Alkire & Foster, 2011).

Existen muchos estudios a nivel econométrico como los modelos de *variable dependiente limitada*, (Probit y Logit) de determinantes de pobreza, que para el caso colombiano se destacan estudioos como los de Núñez y Ramírez (2002) quienes estimaron un modelo de tipo Probit, teniendo en cuenta variables socioeconómicas de los hogares colombianos para los años 1991, 1995 y 2000, dentro de sus hallazgos encontraron que la variable número de personas en el hogar incrementa el riesgo de pobreza para los diferentes años y para el caso del año adicional promedio del hogar disminuye la probabilidad de ser pobre. Recientemente Ariza y Retajac (2020) estimaron un modelo Logit para la pobreza urbana en Colombia y encontraron que para el jefe de hogar “cuentapropista” el riesgo de pobreza se incrementa entre un 4% y 10%, para la variable de ingresos del hogar (recepción de remesas) la probabilidad de ser considerado pobre disminuye entre un 6% y 11%.

La revolución del análisis de datos avanzado (inteligencia artificial, big data y el aprendizaje automático o estadístico) a puesto de manifiesto la discusión de las nuevas formas de identificación y medición de la pobreza a nivel mundial, dentro de las investigaciones más interesantes resaltan las de Blumenstock et al. (2015); Steele et al. (2017) y Chi et al. (2022).

Para el caso de Blumenstock et al. (2015), analizaron datos anonimizados de teléfonos celulares (encuestas de hogares) de Ruanda para predecir la pobreza y la riqueza entrenando un algoritmo de aprendizaje automático. Uno de los mayores expertos en temas sociales (pobreza y desigualdad) y de big data-aprendizaje automático (por sus siglas en inglés machine learning) en la región de América Latina Walter Sosa Escudero ha escrito investigaciones sobre el uso de esta potente herramienta para el entendimiento de la pobreza desde otro enfoque diferente al convencional (econométrico), se destacan estudios como Sosa-Escudero (2018) donde explica las diferentes herramientas del big data y el machine learning para economistas y cómo usarlas para el entendimiento de problemas prácticos como la pobreza, así mismo Sosa-Escudero et al. (2022) estudian a profundidad la aplicabilidad del aprendizaje automático a los estudios de la pobreza, desigualdad y estudios de desarrollo desde un enfoque teórico y práctico y el cómo el usos de estas novedosas y avanzadas técnicas pueden a ayudar al investigador y hacedor de políticas públicas a entender estos temas sociales de gran relevancia para la región de América Latina.

## Datos

### ETL de los datos

### Estadísticas descriptivas

**Figura 1** *Distribución de la pobreza por sexo del jefe de hogar*

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

*Nota.* Elaboración propia

**Figura 2** *Distribución de la pobreza por número de hijos*

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

*Nota.* Elaboración propia

**Figura 3** *Distribución de la pobreza por composición del hogar*

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

*Nota.* Elaboración propia

**Figura 4** *Distribución del ingreso per cápita de la unidad de gasto*

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

*Nota.* Elaboración propia

## Modelos y resultados

### Selección del modelo y entrenamiento

### Optimización de hiperparametros

### Análisis comparativo

### Importancia de los atributos

## Conclusión

1. Grupo 5: Ana María Rojas, Juan Andrés Rincón, Julián Delgado Gutiérrez y Mario Andrés Mercado [↑](#footnote-ref-1)
2. Véase Gasparini et al. (2013) un compendio completo sobre la pobreza y desigualdad tanto desde un punto de vista conceptual, teórico y empírico aplicado en América Latina [↑](#footnote-ref-2)
3. Lora y Prada (2023) en el capítulo 4 de su destacada obra, analiza en términos teóricos las diferentes mediciones del desarrollo humano, que va desde la pobreza monetaria hasta indicadores de desigualdad. [↑](#footnote-ref-3)