

# **Análisis Multidimensional de la Pobreza en Colombia a Través de Datos de Vivienda**

Laura Peralta, Juan David Roldan, Jairo Rueda y Diego Garcia  
Universidad de los Andes  
Maestría en Inteligencia Analítica de Datos  
Septiembre 4, 2023

## **Resumen**

Este proyecto aborda el desafío de identificar la pobreza multidimensional en Colombia utilizando datos de vivienda recopilados en la Encuesta Nacional de Calidad de Vida del DANE en 2022. El índice de pobreza multidimensional es crucial para comprender la situación socioeconómica de las familias, ya que abarca aspectos como educación, salud y estándares de vida. Según estimaciones de la ONU, más de 1.2 mil millones de personas en países en desarrollo viven en pobreza multidimensional. Colombia no es una excepción, y el informe del DANE muestra una disminución significativa en la incidencia de pobreza multidimensional en 2022 en comparación con el año anterior.

El objetivo principal de este estudio es utilizar técnicas de aprendizaje no supervisado y reducción de dimensionalidad para identificar patrones relacionados con la pobreza multidimensional en los datos de vivienda. Esto ayudará a clasificar las viviendas en niveles de riesgo social. En esta etapa se considera usar algoritmos de aprendizaje no supervisado, como K-means y DBSCAN, para agrupar viviendas con características similares. Además, se aplicarán técnicas de reducción de dimensionalidad, como PCA y SVD, para simplificar los datos sin perder información relevante.

Este informe proporciona una revisión de la literatura sobre el Índice Multidimensional de Pobreza y el Índice de Vivienda Adecuada, dos elementos fundamentales en este proyecto. Estos índices son esenciales para medir la pobreza multidimensional y definir criterios de vivienda adecuada. Esta primera fase del proyecto busca sentar las bases para un análisis detallado que contribuirá a comprender mejor la pobreza multidimensional en Colombia y permitirá la identificación de grupos de viviendas en riesgo social. El uso de estas técnicas de análisis de datos es esencial para abordar este desafío y proporcionar información valiosa para la toma de decisiones y políticas públicas.

## **1. Introducción**

Según estimaciones de la ONU para el año 2022 alrededor de 1200 millones de personas residentes de países en vía de desarrollo se encuentra en estado de pobreza multidimensional, estas cifras fueron calculadas en un esfuerzo conjunto entre el programa de desarrollo de las Naciones Unidas (UNDP) y la Iniciativa de Pobreza y Desarrollo Humano de la Universidad de Oxford.

El índice de pobreza multidimensional es una medida en la cual se engloban las necesidades y las faltas de estas en diferentes ámbitos agregados a la educación salud y estándares de vida. Este índice nos ayuda de acuerdo con los criterios contenidos, a determinar la incidencia de la pobreza, así mismo como su intensidad, adicionalmente, también nos ayuda a crear un criterio de comparación no solo entre la misma población estudiada en un país, región si no que los resultados son comparables entre diferentes regiones del mundo. Dentro de los ámbitos tenidos en cuenta mencionados anteriormente, vamos a encontrar que estos componentes se subdividen en criterios específicos presentados a continuación:

1. Educación, compuesto por los años de escolaridad y asistencia al colegio
2. Salud, compuesto por nutrición y mortalidad infantil
3. Estándares de Vida, compuesto por combustible para cocinar, Vivienda, sanidad, agua potable, electricidad y bienes.

En Colombia para el mayo del 2023 el Departamento Administrativo Nacional de Estadísticas (DANE) público el boletín técnico de pobreza multidimensional en Colombia del año 2022, en el cual nos muestra como el porcentaje de pobreza multidimensional se redujo alrededor de 300 puntos básicos comparado con el año inmediatamente anterior.

El Boletín nos muestra como la incidencia de pobreza se redujo 3.1% en general en el territorio, pero específicamente el 3.8% en centros poblados y rural disperso y un 2.8 en cabeceras. Respecto a las regiones, la incidencia más alta de pobreza se registró en la zona Caribe y en la Pacífica (Sin Valle del Cauca), con una incidencia del 21.4% y 20.7% respectivamente.

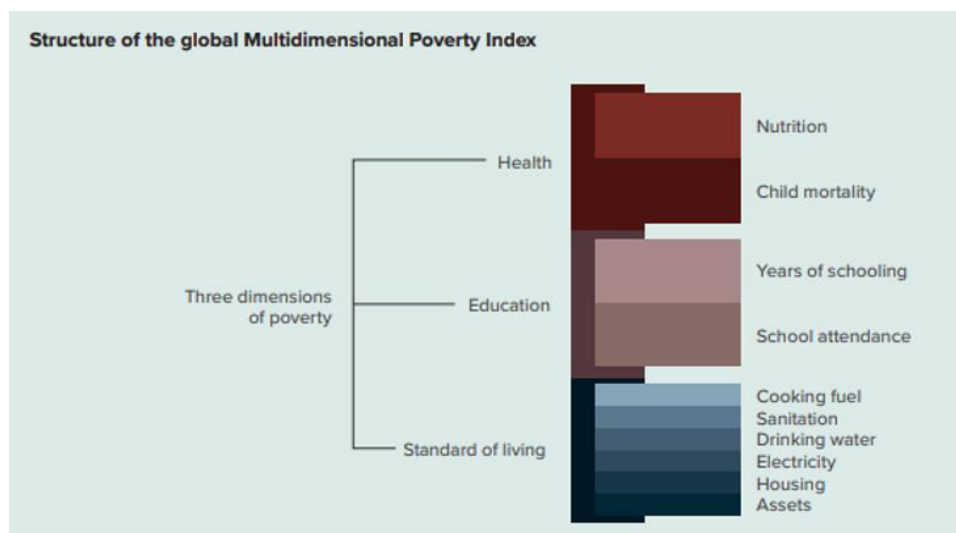
Respecto al aporte en la incidencia entre los componentes del índice, los que más aportaron al índice del 2022 fue la educación con un 35%, y de tercer lugar podemos encontrar las condiciones de Vivienda con un 13.4%.

Lo que nos lleva preguntarnos si a partir del aprendizaje no supervisado y la encuesta nacional de calidad de vida del DANE para Colombia realizada en 2022 específicamente de la data recolectada de datos de vivienda, si es posible identificar si en la población nacional y basados en el índice de vivienda adecuada (Adequate Housing index) propuesto por el banco mundial, se identifican componentes o clústeres específicos que sean relacionados a estimaciones de nivel de pobreza multidimensional en el territorio Colombiano.

## 2. Revisión preliminar de antecedentes en la literatura.

Índice Multidimensional de pobreza: Este es un índice de estimación de pobreza introducido desde el año 2010, el cual es elaborado por parte del Programa de las Naciones Unidas para el Desarrollo, (PNUD en sus siglas en inglés) con colaboración de la iniciativa de Pobreza y Desarrollo Humano de Oxford. Este índice entro a sustituir los índices de pobreza humana. Principalmente este índice de pobreza construye el perfil de cada hogar y las personas que lo habitan basado en 10 criterios específicos contenidos en tres grandes aspectos, educación, salud y estándares de vida, para poder así calcular dos valores específicos: 1. La incidencia de la pobreza o en otras palabras la proporción de las personas que viven en pobreza, y adicionalmente 2. La intensidad de la pobreza o el promedio del puntaje entre las personas catalogadas como pobres. Este índice se ubica entre 0 y 1 en el cual a mayores valores mayor estado de pobreza. Adicionalmente, estos valores con ajustados con inputs específicos para el cálculo en cada país.

Componentes específicos del índice



**Figura 1.** Estructura del índice de pobreza multidimensional global. Adaptado de "Multidimensional Poverty Index" por el Programa de las Naciones Unidas para el Desarrollo (PNUD) y la Iniciativa de Pobreza y Desarrollo Humano de Oxford (OPHI), 2021

DIMENSIONS OF POVERTY	INDICATOR	DEPRIVED IF LIVING IN A HOUSEHOLD WHERE...	WEIGHT	SDG AREA
Health (1/3)	Nutrition	Any person under 70 years of age for whom there is nutritional information is <b>undernourished</b> .	1/6	SDG 2: Zero Hunger
	Child mortality	A child <b>under 18</b> has <b>died</b> in the household in the five-year period preceding the survey.	1/6	SDG 3: Health and Well-being
Education (1/3)	Years of schooling	<b>No</b> eligible household member has completed <b>six years of schooling</b> .	1/6	SDG 4: Quality Education
	School attendance	Any school-aged child is <b>not attending</b> school <b>up to</b> the age at which he/she would complete <b>class 8</b> .	1/6	SDG 4: Quality Education
Living Standards (1/3)	Cooking fuel	A household cooks using <b>solid fuel</b> , such as dung, agricultural crop, shrubs, wood, charcoal, or coal.	1/18	SDG 7: Affordable and Clean Energy
	Sanitation	The household has <b>unimproved</b> or <b>no</b> sanitation <b>facility</b> or it is improved but <b>shared</b> with other households.	1/18	SDG 6: Clean Water and Sanitation
	Drinking water	The household's source of <b>drinking water</b> is <b>not safe</b> or safe drinking water is a <b>30-minute</b> or <b>longer walk</b> from home, roundtrip.	1/18	SDG 6: Clean Water and Sanitation
	Electricity	The household has <b>no electricity</b> .	1/18	SDG 7: Affordable and Clean Energy
	Housing	The household has <b>inadequate</b> housing materials in <b>any</b> of the three components: <b>floor</b> , <b>roof</b> , or <b>walls</b> .	1/18	SDG 11: Sustainable Cities and Communities
	Assets	The household does <b>not own more than one</b> of these <b>assets</b> : radio, TV, telephone, computer, animal cart, bicycle, motorbike, or refrigerator, and does not own a car or truck.	1/18	SDG 1: No Poverty

**Tabla 1.** Dimensiones, indicadores, umbrales y ponderaciones del índice de pobreza multidimensional global. Adaptado de "Multidimensional Poverty Index" por el Programa de las Naciones Unidas para el Desarrollo (PNUD) y la Iniciativa de Pobreza y Desarrollo Humano de Oxford (OPHI), 2021.

Índice de vivienda Adecuada, este índice fue introducido por el banco mundial en el cual se estima cuales debería ser los criterios para considerarse una vivienda adecuada, de acuerdo con esto, para calcular dicho índice se agregan siete criterios para el cálculo total.

El primer criterio introducido es el de acceso al agua, con concordancia al programa conjunto de monitoreo de la OMS y de la UNICEF(JMP), donde se define las fuentes de agua potable las cuales tiene el potencial de entregar agua saludable (JMP 2020). El segundo criterio considerado es acceso a sanidad, dentro de este criterio se encuentra por ejemplo baños con la suficiente estructura para la descarga de los desechos. El tercer criterio está definido como un adecuado espacio para vivir, que básicamente lo podemos resumir que en una vivienda no habiten más de las personas para la cual fue diseñada.

Como quinto criterio tenemos es la seguridad de la tenencia de la vivienda contra situaciones como el desalojo forzoso, y otras amenazas. Como sexto factor encontramos el acceso a la electricidad siguiendo la definición de la agencia internacional de energía que se reduce a tener suficiente acceso a la energía necesaria para alimentar focos de luz, cargadores de celulares, etc. Finalmente, como el ultimo componente, el acceso a cocinar con combustible limpios, tales como electricidad y gas diferenciado de combustibles sólidos como madera, carbón o kerosene.

Adicionalmente este indicador también puede ser ajustado dependiendo de la locación de la vivienda.

### 3. Descripción detallada de los datos

El análisis de variables en este estudio se divide en tres tipos: binarias, categóricas numéricas y de texto. Comenzando con las variables binarias, se utilizan estadísticas descriptivas como la proporción de unos y ceros para cada característica. Se observa que la mayoría de las respuestas son positivas,

con más del 80% en la mayoría de los casos. Se generan gráficas que muestran desbalances en las clases, y se evalúa la relación lineal entre las variables binarias, encontrando una falta de predominancia en las relaciones lineales.

Luego, se aborda el análisis de variables categóricas numéricas, donde se destacan estadísticas descriptivas para variables categóricas ordinales, como la cantidad promedio de familias por hogar. Sin embargo, las variables categóricas no ordinales no brindan información relevante en forma numérica. Se crean gráficas que muestran desbalances en las categorías. En cuanto a la relación lineal entre variables categóricas, se utiliza una matriz de correlaciones, un mapa de calor y un correlograma. Se observa una falta de predominancia de relaciones lineales, aunque se identifican algunas relaciones débiles. No se realiza análisis de la variable de texto debido a la falta de resultados relevantes. En un último gráfico, se resume la relación lineal entre todas las variables, sin encontrar hallazgos significativos. Para obtener más detalles y ver los gráficos, diríjase al notebook de análisis descriptivo.

#### 4. Propuesta Metodológica

Con el objetivo de identificar, componentes o clústeres específicos que sean relacionados a estimaciones de nivel de pobreza multidimensional en el territorio colombiano, basándonos en la encuesta nacional de calidad de vida del DANE para Colombia realizada en 2022, específicamente en los datos recopilados por la encuesta "Datos de la Vivienda" que recopila los resultados de la encuesta realizada a 93,161 hogares colombianos, proponemos la implementación de una metodología robusta que combine técnicas de reducción de dimensionalidad y aprendizaje no supervisado. Esta iniciativa tiene como finalidad abordar el desafío de clasificar los niveles de riesgo social a los que se enfrentan estas familias, utilizando como base los factores socioeconómicos y de calidad de vida contemplados en la encuesta.

##### **Algoritmos de aprendizaje no supervisado:**

Los algoritmos de aprendizaje no supervisado se utilizan para identificar patrones en datos que no están etiquetados. Es decir, los datos no contienen información sobre la clase o categoría a la que pertenecen los puntos de datos.

En este caso, los datos no están etiquetados con información sobre el nivel de pobreza multidimensional de las viviendas. Por lo tanto, se requiere un algoritmo de aprendizaje no supervisado para identificar grupos de viviendas con características similares que puedan ser relacionadas con el nivel de pobreza multidimensional, algunos algoritmos de aprendizaje no supervisado que son candidatos a utilizarse para este problema son:

**Algoritmos basados en centroides:** Estos algoritmos son adecuados para identificar grupos en los datos donde cada grupo se representa mediante un centroide, que es el promedio de todos los puntos del grupo.

- **K-means:** Este algoritmo divide los datos en k grupos, de manera que la suma de las distancias entre los puntos de datos y el centroide de su grupo sea mínima.
- **K-median:** Este algoritmo es similar a k-medias, pero utiliza las medianas de los datos en lugar de los centroides.

**Algoritmos basados en densidades:** La elección de DBSCAN se justifica por su capacidad para identificar grupos en función de áreas de alta densidad de datos. Esto es fundamental, ya que ciertas

áreas geográficas pueden tener características socioeconómicas similares, y DBSCAN es capaz de detectar estos patrones sin asumir la forma y el tamaño de los grupos de antemano.

### **Reducción de la Dimensionalidad:**

Para gestionar la complejidad inherente de los datos recopilados por la encuesta del DANE y mejorar la eficiencia computacional, se propone la utilización de técnicas de reducción de dimensionalidad. Estas técnicas son esenciales para transformar los datos originales en representaciones de menor dimensión, lo que no solo facilita el análisis, sino que también contribuye a preservar la información relevante para la clasificación del nivel de riesgo de las viviendas.

Dado el contexto del problema y la organización, consideramos que las técnicas de Análisis de Componentes Principales (PCA) y Descomposición en Valores Singulares (SVD) son las más apropiadas. Estas técnicas son ampliamente reconocidas y utilizadas en análisis de datos, y permiten reducir la dimensionalidad mientras mantienen la información esencial. Además, la interpretabilidad de los resultados se ve mejorada, lo que facilita la toma de decisiones basada en los hallazgos.

## **5. Bibliografía**

- Organización de las Naciones Unidas (ONU). (2022). Informe sobre Desarrollo Humano 2022. Nueva York: ONU.
- Departamento Administrativo Nacional de Estadísticas (DANE). (2023, mayo). Boletín Técnico de Pobreza Multidimensional en Colombia 2022. Bogotá, Colombia: DANE.
- UNDP: Programa de las Naciones Unidas para el Desarrollo. (2022). Informe sobre Desarrollo Humano 2022. Nueva York: Autor.
- Iniciativa de Pobreza y Desarrollo Humano de la Universidad de Oxford: Oxford Poverty & Human Development Initiative. (2022). MPI Database. Oxford: Autor.
- PNUD: Programa de las Naciones Unidas para el Desarrollo. (2010). Manual de medición de la pobreza multidimensional. Nueva York: Autor.
- Banco Mundial: Banco Mundial. (2005). Adequate Housing: Concepts, indicators, and policies. Washington, DC: Autor.