

# Identificación de la pobreza en Colombia: comparación entre el índice de pobreza multidimensional y el sistema de identificación de beneficiarios de programas sociales

## Poverty identification in Colombia: Comparison between the Multidimensional Poverty Index and the Beneficiaries Identification Social System Programs

### Resumen

El estudio examina la identificación de la pobreza en Colombia, comparando el índice de pobreza multidimensional y el sistema de identificación de beneficiarios de programas sociales. Se analiza los resultados para Colombia, en Medellín, por medio de indicadores de concordancia y una tabla de contingencia. Se utiliza un modelo de aprendizaje estadístico para estimar la clasificación del sistema de identificación de beneficiarios de programas sociales, con una exactitud balanceada de 78%. Se realiza un análisis cualitativo que revela la pobreza multidimensional de cuatro familias no consideradas pobres por el sistema de identificación. Estos hallazgos resaltan una baja e insignificante concordancia entre ambas medidas. Los resultados contribuyen a mejorar la definición y la identificación de la pobreza en el país, beneficiando la política pública.

**Palabras clave:** programas sociales, indicadores de pobreza, análisis de datos, política pública, Medellín (Colombia).

**Clasificación JEL:** I38, I32, C55, I38, D63.

### Abstract

This study examines the identification of poverty in Colombia by comparing the multidimensional poverty index and the social program beneficiary identification system. The results are analyzed in Colombia and the city of Medellín using measures of agreement and a contingency table. A statistical learning model is employed to estimate the classification of the social program beneficiary identification system, achieving a balanced accuracy of 78%. Qualitative analysis reveals the multidimensional poverty of four families not identified as poor by the beneficiary identification system. These findings highlight a low and insignificant agreement between the two measuring systems. The results contribute to improving the definition and identification of poverty in the country, supporting public policy design.

**Keywords:** Social programs, poverty indicators, data analysis, public policy, Medellín (Colombia).

**JEL Classification:** I38, I32, C55, I38, D63.

<sup>a</sup> Instituto de Estudios Regionales, Universidad de Antioquia, Medellín, Colombia. Correo electrónico: guberney.muneton@udea.edu.co

## Introducción

La pandemia de Covid-19 exacerbó las condiciones de pobreza en el mundo, incluyendo a Colombia. Según el Banco Mundial (BM, 2022), la tasa de pobreza extrema en Colombia aumentó del 8.4 en 2019 a 9.3% en 2020, lo que significa que el número total de personas que vivían con menos de 2.15 dólares estadounidenses al día aumentó de 648 millones a 719 millones en 2020. El aumento de la pobreza en países pobres refleja economías con más informalidad, sistemas de protección social débiles y sistemas financieros poco desarrollados.

En una evaluación no monetaria en 111 países, el índice global de pobreza multidimensional de 2022 reveló que el 19.1% se encuentra en pobreza multidimensional (UNDP, 2022). Se estima que el progreso en la reducción de la pobreza multidimensional se ha retrocedido cerca de 10 años debido a la pandemia (UNDP, 2022). Tanto el informe del BM como el informe del PNUD coinciden en la importancia de la política fiscal para activar la recuperación hacia la senda de reducción de la pobreza, con miras a alcanzar los objetivos de desarrollo sostenible (ODS) para 2030.

En Colombia, según datos oficiales del Departamento Nacional de Estadística (Dane), la línea nacional de pobreza extrema per cápita en 2021 fue de 161 099 pesos al mes, mientras que la línea nacional de pobreza per cápita fue de 354 031 pesos al mes. En 2021, el porcentaje de población en pobreza extrema fue del 12.2%, mientras que, en pobreza, fue del 39.3%. Colombia utiliza, principalmente, tres formas para identificar a la población pobre: la pobreza monetaria, la pobreza multidimensional y el Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales — Sisben (DNP, 2012).

Para la pobreza monetaria, el Dane calcula los valores de la línea de pobreza y pobreza extrema del país, así como sus áreas espaciales, como referentes para comparar los ingresos de las personas. La línea de pobreza extrema representa el costo mensual per cápita necesario para adquirir una canasta básica de alimentos, mientras que la línea de pobreza es el costo mensual

per cápita para adquirir bienes y servicios básicos, además de los alimentos. Por su parte, el índice de pobreza multidimensional (IPM), definido para Colombia, sigue la metodología Alkire-Foster, considerando pobre al hogar que se encuentre, simultáneamente, privado en más del 33% de los indicadores ponderados definidos para la medida (Angulo-Díaz y Pardo, 2016).

El Sisben es otra medida utilizada por el Estado colombiano para focalizar la ayuda social (DNP, 2016). Es un tipo de medida del bienestar, que clasifica a la población en pobres, vulnerables y no pobres y está compuesto por una encuesta y un sistema estadístico de clasificación. Desde 2019, se aplica la cuarta versión del Sisben, creado en 1995, y su objetivo principal es combinar las dimensiones de la pobreza multidimensional y los ingresos como principales aspectos para la identificación de potenciales beneficiarios de la política social. El Sisben propone un modelo que aproxima la capacidad de la población de generar ingresos, buscando que el resultado sea similar al comportamiento de la pobreza monetaria y multidimensional de Colombia (DNP, 2016).

Con ello en mente, en el artículo se analiza la clasificación del Sisben en su última versión (IV), comparándola con el índice de pobreza multidimensional de Colombia (Angulo *et al.*, 2016). Para el análisis, se utilizó el resultado de la clasificación del Sisben IV como variable exógena; y se incluyeron experimentos donde se cambian los pesos de las variables que componen el indicador de pobreza multidimensional. El análisis se realizó solo con la pobreza multidimensional, porque la base de datos del Sisben IV no contiene suficiente información para la estimación de la pobreza monetaria. Esa falta de información sobre ingresos y gastos del hogar impide realizar los cálculos de los indicadores sobre las mismas personas.

Es relevante comprender la versión IV del Sisben, porque recoge dos elementos interesantes para la política pública: (1) su implementación supone un cambio tanto teórico como metodológico respecto a la versión III y (2) propone el objetivo de clasificar a las personas teniendo en cuenta el comportamiento de la pobreza monetaria y multidimensional. Dentro de este segundo

<sup>b</sup> Universidad de Antioquia. Medellín, Colombia. Correo electrónico: caaperezag@unal.edu.co

<sup>c</sup> GITA Lab., Universidad de Antioquia, Medellín, Colombia. Correo electrónico: rafael.orozco@udea.edu.co

elemento, la medición de la capacidad del Sisben IV para integrar a la población multidimensionalmente pobre en su resultado es parte de las respuestas que se ofrecen en el presente artículo.

Las medidas identifican aspectos complementarios del mismo fenómeno, lo que permite un mayor conocimiento del problema; también plantea un desafío para la focalización de la política fiscal (Villatoro y Santos, 2019). La identificación de la población pobre es crucial para la implementación efectiva de políticas sociales; y minimizar los errores de exclusión e inclusión es fundamental para mejorar la precisión de estas medidas (Sen, 1992). Al identificar la población excluida/incluida de la medición del Sisben IV y, al mismo tiempo de la medición de pobreza multidimensional, interesa analizar el comportamiento de los grupos y las variables de importancia para la clasificación.

En el artículo también se analizan las variables que conforman las medidas y se estima la clasificación que realiza el Sisben IV, para identificar, de ese modo, las variables que tienen mayor aporte a dicha clasificación. Es necesario realizar la estimación, porque el documento oficial del Sisben IV solo enuncia las variables independientes como agregados, es decir, no publica los indicadores específicos que el Gobierno utiliza para focalizar la política social.

Para esta estimación, se aplica el algoritmo de árboles aleatorios (random forest), para analizar las variables de importancia en la asignación de los grupos. El algoritmo de aprendizaje estadístico se utiliza para revelar las variables que más aportan a la clasificación del Sisben. Esto permite realizar un análisis transparente y justo, con respecto a las mediciones de pobreza, lo cual incentiva el control social para evitar los daños sociales que pueden quedar ocultos, cuando se utilizan metodologías tipo caja negra (Corbett-Davies y Goel, 2018). La justificación de mantener en secreto estadístico las variables específicas de la clasificación del Sisben, argumentando la estabilidad macroeconómica, plantea dudas sobre la transparencia y la justicia de estas medidas (DNP, 2016).

Adicionalmente, es importante realizar ejercicios de sensibilidad de las medidas de pobreza, a fin de ampliar la discusión sobre un concepto de pobreza que incluya a las diferentes personas identificadas como pobres

por las diversas mediciones. La capacidad de estos algoritmos para clasificar correctamente a la población pobre depende de su diseño y de los datos. Ello hace necesario evaluarlos continuamente para minimizar errores y mejorar la precisión en la identificación de beneficiarios de programas sociales (Villatoro y Santos, 2019). El artículo se estructura de la siguiente manera: en la sección I, se presentan los datos y los métodos de análisis. En la sección II, se presentan los resultados. La sección III plantea una discusión sobre los hallazgos. Por último, en la sección IV se incluyen las conclusiones derivadas del estudio, así como sus limitaciones.

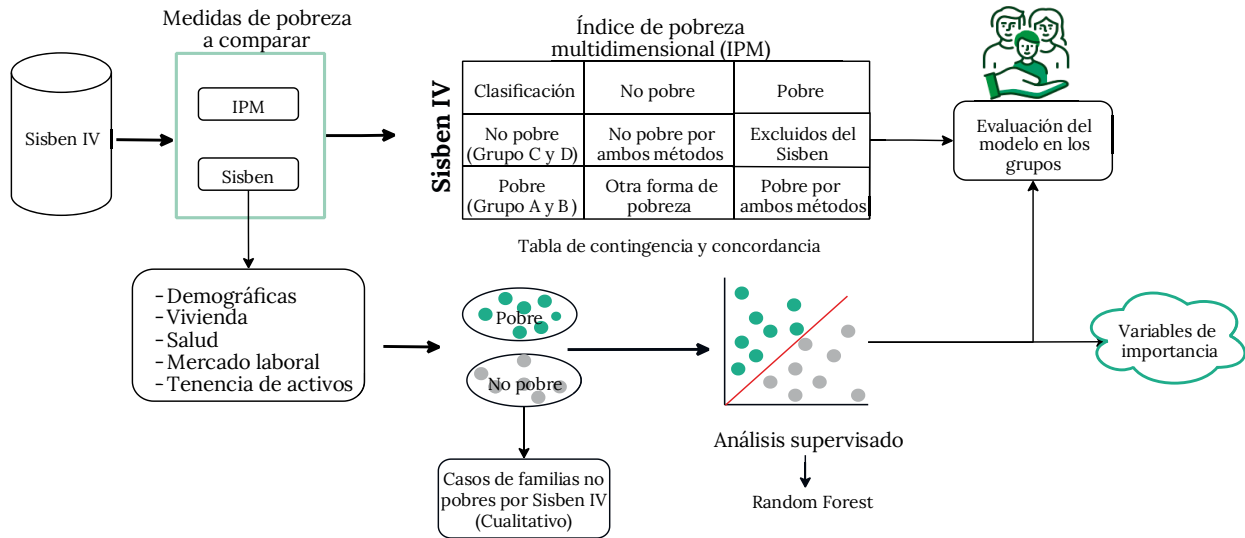
## I. Metodología

La metodología se muestra en la figura 1. Se analizaron dos casos: Colombia y Medellín. Se utilizó la base de datos del Sisben IV para ambos. Se generaron tres versiones del IPM para clasificar a la población como pobre o no pobre. Se comparó esta clasificación con la del Sisben IV, donde los grupos A y B se consideran pobres y los grupos B y C, no pobres.

Asimismo, se realizó la comparación, utilizando tablas de contingencia e indicadores de concordancia. Luego, se empleó un modelo de árboles aleatorios para identificar las variables más importantes en la clasificación del Sisben. Se analizaron los grupos excluidos por el Sisben, pero considerados pobres por el IPM; así como los pobres identificados por ambos métodos. Como parte de la metodología, se realizó un análisis cualitativo de cuatro familias, clasificadas como no pobres por el Sisben IV, pero con evidencia suficiente de pobreza multidimensional.

### A. Bases de datos

Se utilizaron dos bases de datos, proporcionadas ambas por entes gubernamentales. La primera fue la base del Sisben IV de marzo de 2021 para todo Colombia, disponible en la página de datos abiertos del DNP: <https://anda.dnp.gov.co/index.php/catalog/135>. La segunda fue la del Sisben IV para Medellín, con corte a abril de 2022. Esta base de datos se obtuvo mediante un derecho de petición a la Alcaldía de Medellín. De acuerdo con el DNP, la base de Colombia es una muestra anonimizada, con cobertura geográfica hasta nivel de municipio. Ambas bases del Sisben se construyeron



**Figura 1.** Metodología general del análisis

Fuente: elaboración propia.

por demanda de los hogares y con barridos territoriales. La base de datos de Medellín contiene todos los registros que hasta la fecha (abril del 2022) tenían el cálculo certificado del DNP. Por tanto, la base de Medellín tiene más registros del municipio que los contenidos en la base de Colombia.

En relación con los datos, es importante mencionar algunas dificultades con la base para Colombia. En el repositorio del DNP, se encuentran tres tablas asociadas a la base de datos: vivienda, conformación de hogares y personas. Esta última tabla contiene el cálculo aproximado del índice de pobreza multidimensional para cada individuo; y precisamente este dato es el utilizado aquí para realizar la comparación entre el IPM y el Sisben IV. Sin embargo, el repositorio tiene problemas de especificación que no permiten unir las bases de datos. Una de las razones principales de ello es que el Gobierno define de manera independiente el concepto de hogar y vivienda. En una vivienda pueden convivir varios hogares. Esta diferencia conceptual genera un problema de etiquetado de los datos.

Adicionalmente, al observar los datos proporcionados por el gobierno se encuentra que la base de personas (a nivel de país) tiene 3706298 registros, mientras que la base de datos de hogares contiene 1341695 registros; y solo 3049 valores únicos que los etiqueta como

consecutivo de vivienda. Es decir, los registros de la base de datos de hogar y de personas se distribuyeron solamente en 3046 viviendas.

Según estos datos, una vivienda en Colombia albergaría un promedio de 440 hogares y 1216 personas, lo que a todas luces es un error. Por tanto, los datos de hogar, vivienda y personas no se pueden unir. Esto impide contar con una base de datos completa, con datos correctamente etiquetados que permitan entrenar un modelo de clasificación. Lo anterior no altera el análisis de la concordancia entre las mediciones de pobreza, sino que solamente afecta al proceso de modelado que permite evidenciar las variables de importancia de la clasificación del resultado del Sisben.

Por su parte, la base de datos de la ciudad de Medellín tiene coherencia para la integración de las tablas de viviendas, hogares y personas. Sin embargo, no cuenta con el cálculo del IPM, por tanto, este índice fue calculado por los autores, siguiendo los lineamientos del DNP y la medición aceptada para Colombia (Angulo *et al.*, 2016).

## B. Características de las bases de datos

En la base de Sisben IV para Colombia se reportaron 3706298 personas, lo cual corresponde a 1341695 hogares, es decir, alrededor de tres personas por hogar.

La distribución por sexo es equilibrada, el 50.84 % corresponde a mujeres y el 49.15 %, a hombres. La mayor proporción de personas (36%) se encuentra en el rango de 29-60 años de edad. De los 1104 municipios de Colombia, la base de datos tiene representación de 1100 (99.63%).

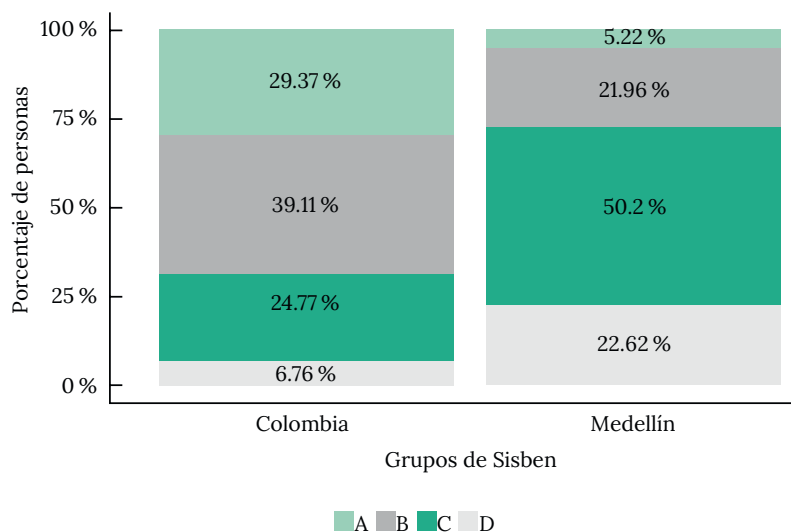
La distribución de la población en los grupos del Sisben IV se observa en la figura 2. El 29.37% se encuentra en pobreza extrema (grupo A); en el siguiente grupo, el B, se encuentra el 39.11% que estarían en pobreza moderada. Con menores porcentajes, el grupo C (vulnerables) y D (ni pobres ni vulnerables), con el 24.7% y 6.76 %, respectivamente. Por tanto, sumando A y B, se encuentra el 68.48% que se considera pobre, mientras que en los grupos C y D, el 31.52% se clasifica como no pobre. Aunque hay una distribución equilibrada entre hombres y mujeres, en la clasificación por sexo, hay más mujeres que hombres pobres (Grupo A). Al filtrar la base por jefes de hogar, se observa una diferencia de 10 puntos porcentuales entre las mujeres y los hombres que forman parte del grupo A (27.97% mujeres frente a 17.52% hombres).

En la base de datos del Sisben de Medellín, con corte a abril de 2022, se encontraban registradas 1331472 personas, distribuidas en 369079 hogares, con una mediana de tres personas por hogar. De las personas registradas, el 45.4% corresponde a mujeres y el restante 54.6% a hombres. El promedio de edad de

las personas es 35 años. La base de datos contiene personas de los diferentes grupos poblacionales y de 17 territorialidades del municipio (comunales y corregimientos) de las 21 con las que el municipio cuenta; por alguna razón no aparecen en la base de datos las áreas rurales.

La distribución de la población en los grupos del Sisben IV se observa en la figura 3. En los dos primeros grupos, A y B, donde se clasifican los pobres extremos y los pobres moderados, se encuentra el 27.18 % de las personas (5.22% pobres extremos y 21.96 % pobres moderados). En los otros dos grupos, C y D, clasificados como vulnerables y no pobres, se encuentra el 72.82% de las personas de la base de datos.

Al observar la clasificación del Sisben de acuerdo con el sexo de quien se reconoce como jefe de hogar, se evidencia que una mayor proporción de mujeres cabeza de hogar se encuentra clasificada en los grupos A y B en comparación con los hombres que ejercen el mismo rol. Al parecer, el Sisben de Medellín evidencia una especie de feminización de la pobreza en las mujeres consideradas jefes de hogar. Esta misma tendencia se evidencia al cruzar la variable del nivel educativo alcanzado por la jefa del hogar, ya que a medida que aumenta el nivel educativo se observa una disminución de la pobreza. El nivel educativo con mayor proporción de población es la educación básica primaria.



**Figura 2.** Población en los grupos del Sisben IV, Colombia con corte al 2021 y Medellín al 2022.

*Fuente:* elaboración propia con base en datos oficiales de Colombia y Medellín.

## II. Métodos

### A. Clasificación del Sisben IV

El Sisben IV utiliza un enfoque relativo para clasificar a las familias. Este enfoque se basa en una regresión cuantílica que asigna un número de orden a cada familia, en términos de su capacidad para generar ingresos. La posición de cada familia depende de su situación relativa, en comparación con otras familias en el mismo momento. Este método permite identificar a los hogares con mayor necesidad de asistencia social, con base en su capacidad relativa para generar ingresos.

El modelo considera diferentes aspectos, como los demográficos, mercado laboral, vivienda, tenencia de activos y salud. Aunque estos aspectos son mencionados en el documento Conpes 3877 (DNP, 2016), no se proporciona información detallada sobre las variables específicas utilizadas para el cálculo. El Gobierno nacional argumenta que la divulgación de estas variables podría poner en riesgo la estabilidad macroeconómica del país, por lo que se mantiene oculto el algoritmo para evitar posibles fraudes por parte de los ciudadanos.

El texto oficial también omite la discusión sobre la selección de variables y sus pesos, a pesar de que esto ha sido objeto de debates académicos y sociales en la implementación de indicadores de bienestar, pobreza o calidad de vida (Alkire *et al.*, 2015; Atkinson, 1987; Decancq y Lugo, 2013; Greco, 2018). Aunque la descripción oficial del método menciona que las variables se seleccionan a partir de estudios previos, el documento oficial no proporciona referencias bibliográficas de respaldo de la afirmación. El resultado de la estimación del modelo que usa el Sisben entrega un indicador con un rango de variación entre 0 y 100. El indicador se estima por zona geográfica (cabecera y rural) dentro de los 32 departamentos colombianos. El resultado de la estimación se expresa en cuatro grupos, que el DNP describe de la siguiente manera:

|         |  |
|---------|--|
| Grupo A | Población en extrema pobreza o con menor capacidad de generación de ingresos. Se divide en 5 subgrupos que van de A1 a A5.             |
| Grupo B | Población en pobreza moderada, pero con mayor capacidad de generar ingresos que los del grupo A. Tiene 7 subgrupos, desde B1 hasta B7. |

|         |   |
|---------|---|
| Grupo C | Población vulnerable o en riesgo de caer en condición de pobreza. Se descompone en 18 subgrupos, de C1 a C18. |
|---------|---|

|         |  |
|---------|--|
| Grupo D | Población no pobre ni vulnerable. Tiene 21 subgrupos, de D1 a D21. |
|---------|--|

Según el Gobierno colombiano, partiendo de los significados expresados para los puntajes del Sisben IV, puede identificarse a la población pobre de la siguiente manera: son pobres las personas y los hogares que pertenecen a los grupos A y B; y son no pobres las clasificadas en los grupos C y D. Cabe anotar que la unidad de análisis del Sisben IV es el hogar, el instrumento funciona a partir de fichas de hogares que indagan sobre las condiciones a nivel de hogar e individuo. Sin embargo, el resultado de la clasificación es igual para todos los miembros del hogar. Para los efectos de la comparación de las medidas de pobreza multidimensional, se confirma por el DNP que con la información contenida en la ficha de caracterización socioeconómica que se utiliza para el cálculo del algoritmo del Sisben IV, se puede calcular el índice de pobreza multidimensional (DNP, 2016).

#### 1. Índice de Pobreza Multidimensional (IPM)

El IPM de Colombia sigue el método Alkire-Foster (2011), un enfoque basado en el conteo de doble corte que incorpora los desarrollos teóricos del enfoque de las capacidades (Sen, 1985, 1992, 1999). El método sigue un enfoque absoluto de la pobreza multidimensional, aunque es posible realizar un ordenamiento de los hogares que se evalúan por medio del puntaje de las privaciones ponderadas. Para el método, se definen dos umbrales: *u* no para determinar la privación en cada indicador y otro para determinar de cuántos indicadores es necesario estar privado, para ser considerado pobre multidimensional.

En términos específicos, este método consiste en identificar las privaciones en un conjunto de indicadores, donde cada indicador tiene un umbral definido que debe ser alcanzado para satisfacerlo. En primer lugar, se determina la cantidad de privaciones ponderadas que un hogar experimenta. Luego, se establece un puntaje mínimo de privaciones ponderadas para que un hogar sea considerado pobre. El puntaje de cada hogar se calcula sumando las privaciones

ponderadas por los pesos de los indicadores. En el caso del IPM de Colombia, el umbral de pobreza se establece en un 0,33, lo que coincide con el valor de corte del IPM global (Alkire *et al.*, 2020; UNDP y OPHI, 2021). En Colombia, se han definido quince indicadores agrupados en cinco dimensiones, y cada uno tiene su umbral de privación asociado (DNP, 2012; Angulo *et al.*, 2016). La unidad de análisis del IPM de Colombia es el hogar y, aunque algunas privaciones se evalúan por individuo, el resultado se asigna al hogar.

En resumen, un hogar en Colombia se considera pobre si está privado en más del 33% de los indicadores ponderados. Dicho porcentaje es el umbral de la pobreza ( $k$ ). El conteo de la cantidad de personas que supera ese umbral se llama *tasa de recuento* o incidencia de la pobreza, calculado como  $H_k = q_k/n$ , donde  $q_k$  es el número de hogares identificados como pobres, teniendo en cuenta el umbral  $k$ , con  $n$  que corresponde al número total de hogares evaluados.

Además de la tasa de recuento, el método AF considera otro indicador que evalúa la profundidad de la pobreza o intensidad de la privación. Dicho indicador está definido como el promedio de las privaciones que enfrenta la población identificada como pobre. El indicador de la intensidad está dado por  $A = \sum_{i=1}^n c_i(k)/qk$ , donde  $c_i(k)$  corresponde a las privaciones ponderadas del hogar  $i$  identificado como pobre. Ahora bien, el IPM o tasa de recuento ajustada  $M_0$  se expresa como el producto de dos subíndices: la incidencia  $H_k$  y la intensidad  $A$ . Entonces,  $M_0 = H_k A$  (Alkire y Foster, 2011).

## 2. Tabla de contingencia e índices de concordancia

Las tablas de contingencia, es decir matrices numéricas que representan conteos o frecuencias, se utilizan para identificar relaciones entre variables (Conover, 1999). En el contexto de los estudios sobre la pobreza, estas tablas se han utilizado para evaluar similitudes en la incidencia de la pobreza según diferentes métodos (Boltvinik, 2013; Villatoro y Santos, 2019).

La concordancia, por su parte, permite evaluar las diferencias o similitudes entre dos pares de medidas. Cuando dos medidas de pobreza son concordantes en su totalidad, significa que identifican a las mismas personas como pobres de manera independiente (Viera *et al.*, 2005). Por el contrario, si las medidas muestran

una falta total de concordancia, significa que evalúan aspectos diferentes. Para evaluar la concordancia, se utilizan medidas basadas en la tabla de contingencia.

Una medida general para comparar dos medidas de pobreza es sumar los pobres y no pobres según ambos métodos; y expresar el valor como porcentaje del total, lo cual se conoce como concordancia general. Otra medida, más robusta, de coincidencia es el coeficiente  $k$ , el cual se define de la siguiente manera (Viera *et al.*, 2005):

$$k = \frac{P_0 - P_e}{1 - P_e}$$

Donde  $p_e$  es la proporción de concordancia esperada por azar y  $p_0$  es la proporción de concordancia observada.  $P_e$  y  $P_0$ , se utilizó una tabla de contingencia que compara las clasificaciones del IPM y del Sisben IV (cuadro 1).

**Cuadro 1.** Tabla de contingencia

|                | Pobre (Sisben IV) | No pobre (Sisben IV) |
|----------------|-------------------|----------------------|
| Pobre (IPM)    | $n_1$             | $m_1$                |
| No pobre (IPM) | $n_0$             | $m_0$                |

*Notas.*  $n_1$  es el número de individuos clasificados como pobres tanto por el IPM como por el Sisben IV.  $m_1$  es el número de individuos clasificados como pobres por el IPM pero no por el Sisben IV.  $n_0$  es el número de individuos clasificados como no pobres por el IPM pero sí por el Sisben IV.  $m_0$  es el número de individuos clasificados como no pobres tanto por el IPM como por el Sisben IV.

Fuente: elaboración propia.

Entonces, la proporción de concordancia observada ( $P_0$ ) se calcula de la siguiente manera:

$$P_0 = \frac{n_1 + m_0}{n_1 + m_1 + n_0 + m_0}$$

La proporción de concordancia esperada por azar ( $P_e$ ) se calcula como se indica enseguida:

$$P_e = \left( \frac{(n_1 + m_1) \times (n_1 + n_0)}{(n_1 + m_1 + n_0 + m_0)^2} \right) + \left( \frac{(n_0 + m_0) \times (m_1 + m_0)}{(n_1 + m_1 + n_0 + m_0)^2} \right)$$

Una vez calculados  $P_0$  y  $P_e$ , se sustituyen en la fórmula del índice kappa ( $k$ ) para obtener el valor de la

concordancia ajustada. El índice  $k$  tiene en cuenta el acuerdo que puede darse únicamente por azar (Viera *et al.*, 2005). Por tanto, valores inferiores a 0.2 suelen interpretarse como insignificantes; valores entre 0.2 y 0.4, como concordancia baja; valores entre 0.4 y 0.6 como concordancia moderada; y valores mayores a 0.6 como concordancia alta.

## B. Experimentos

Se realizó un análisis de concordancia entre el IPM y los grupos del Sisben (A, B, C, D), utilizando la base de datos de Colombia y Medellín. En este análisis, se consideraron pobres las personas pertenecientes a los grupos A y B, y no pobres los grupos C y D. Los análisis de los indicadores de concordancia entre la medición de pobreza multidimensional y el Sisben IV se muestran de acuerdo con los diferentes puntos de corte del IPM. Los experimentos descritos a continuación buscan analizar la consistencia entre dos medidas de pobreza: la pobreza según el Sisben IV y la pobreza multidimensional.

*Experimento 1.* Se compararon los IPM oficiales de Colombia y Medellín con el Sisben IV. Para Colombia, se tomó el valor del IPM reportado. Para Medellín, se calculó siguiendo los lineamientos de la metodología oficial de Colombia (Angulo *et al.*, 2016). El cálculo tuvo en cuenta los ajustes del DNP para aproximar algunos indicadores de las tres dimensiones (cuadro 2). Con este experimento, se analizó el IPM oficial con los ajustes pertinentes del DNP.

*Experimento 2.* Tomando como referencia el IPM oficial de Colombia y el calculado para Medellín, se afectaron los pesos de los indicadores, pasando de una ponderación anidada a una definición de pesos iguales para

todos los indicadores. Entonces, se comparó el IPM y el Sisben IV, considerando pesos iguales para los quince indicadores de la medición oficial. En contraste con la ponderación oficial, que considera pesos iguales en las cinco dimensiones y pesos proporcionales a los indicadores que forman parte de la dimensión. El experimento buscó asignar una importancia igual a todas las privaciones, a fin de evaluar si, al mantener una estructura de pesos constante por indicadores, los conjuntos de pobres cambian. Con la modificación, entonces, se analizó la robustez de la medición de pobreza.

## C. Cálculo del IPM a partir de la base de datos del Sisben IV

La base de datos del Sisben IV de Medellín no contenía el cálculo del proxy del IPM. Por consiguiente, fue necesario calcular el índice siguiendo la metodología oficial. Para calcular el IPM a partir de la base de datos del Sisben, se siguió la metodología oficial. Este índice considera pobre un hogar, si está privado en un puntaje de más del 0,33 de los indicadores ponderados definidos para la medida. Las tablas con las dimensiones, indicadores y ponderaciones se encuentran en los Anexos 1 y 2.

## D. Estimación de la clasificación de pobres y no pobres del Sisben IV

Se propone un modelo de clasificación de la población etiquetada como pobre por el Sisben IV, para analizar la capacidad de discriminación, usando las variables contenidas en la base de datos del Sisben IV. Además, el resultado del modelo permitió analizar los grupos de la tabla de contingencia, para ampliar la evaluación de la consistencia de las medidas de pobreza multidimensional (Sisben IV e IPM).

**Cuadro 2.** Ajustes metodológicos para calcular el IMP con el Sisben IV

| Indicador ajustado                              | Diferencia con el indicador oficial   |
|---|---|
| Acceso a servicios cuidado primera infancia     | La ficha Sisben no tiene la pregunta acerca de recibir alimentos en el centro educativo al que se asiste, para los niños de cinco años. |
| Trabajo   | Clasificación de ocupados y desocupados a partir de una sola pregunta de actividad principal en el último mes.                          |
| Servicios públicos y condiciones de la vivienda | En eliminación de excretas para la zona rural, la privación incluye letrina unida con bajamar.  |

Fuente: DNP-Sisben IV.



La estimación también entregó evidencias sobre las variables de mayor importancia para lograr el resultado del modelo. Tener las variables de importancia ayuda a comprender cuáles son las variables principales para la clasificación de pobreza del Sisben IV, algo no evidente en la descripción de la metodología del Gobierno (DNP, 2016). El modelo se estimó únicamente para Medellín, dada la imposibilidad de unir los datos de las personas con los datos de los hogares y la vivienda en la base de datos disponible para toda Colombia, como se detalló en la sección de las bases de datos de la metodología. La estimación de la clasificación de la pobreza usada aquí permite un análisis más profundo de los resultados del Sisben, sin alcance para sugerir alternativas de un nuevo método de clasificación de los pobres por Sisben.

Para la estimación de la clasificación de los pobres, se declararon dos categorías: pobres y no pobres. La primera está compuesta por los grupos A y B del Sisben IV y la segunda, por grupos C y D. La estimación se realizó con un modelo de bosques aleatorios (RF), entrenado para realizar la clasificación bi-clase. Un RF es un algoritmo de aprendizaje estadístico o aprendizaje de máquina (*machine learning*) que ha mostrado desempeño superior a otros algoritmos para la predicción de la pobreza en diversas zonas geográficas (Sohnesen y Stender, 2017; Usmanova et al., 2022a). Una de las principales características del RF es su capacidad para manejar variables de tipo nominal y ordinal, además de la posibilidad de identificar las variables de mayor importancia para la estimación (Schonlau y Zou, 2020).

El algoritmo RF, eficiente en el manejo de grandes volúmenes de datos, puede capturar relaciones complejas entre variables. Una de las ventajas clave del algoritmo es su capacidad para evaluar la importancia de cada variable en la clasificación. Esto permite identificar los factores más determinantes en la asignación de los grupos del Sisben IV.

Este algoritmo constituye un método de ensamblaje que usa diferentes árboles de decisión construidos sobre muestras de entrenamiento. El proceso de elección de variables es aleatorio y entrena diferentes árboles que buscan reducir la varianza (es decir, reducir la incertidumbre). Con el modelo es posible estimar el error de forma interna, la correlación y las

variables más relevantes (Breiman, 2001). En general, es un modelo robusto a *outliers* y ruido. En el contexto de la clasificación de pobreza, donde puede haber variabilidad y datos incompletos, esta característica mejora la fiabilidad del modelo.

Las variables explicativas para el modelo se eligen de acuerdo con las enunciadas en el documento Conpes 3877, donde se define el modelo usado por el Sisben IV (DNP, 2016). Este documento solo enuncia que las variables relacionadas con aspectos demográficos, de la vivienda, la salud, el mercado laboral y la tenencia de activos. Por ello, se evaluó un modelo que agregara todas las variables asociadas disponibles en la base de datos del Sisben IV. Cabe anotar que durante la investigación se hizo la solicitud oficial de las variables completas al DNP de Colombia. La respuesta de dicha oficina fue negativa, argumentando confidencialidad y compromiso de los datos con la estabilidad macroeconómica del país. Por consiguiente, se hizo la estimación bajo el indicio de las variables definidas en el Conpes de referencia.

La estimación de los modelos se realizó con una colección de paquetes para R *project* llamada “tidymodels” (Kuhn y Wickham, 2020), que utiliza principios de *tidyverse*, lo cual permite hacer análisis de datos con una gramática consistente (Wickham et al., 2019). Las estimaciones se ejecutaron en un servidor con 376 GB de memoria RAM y los cálculos tardaron unas 48 horas. Dado que las bases de datos contienen más de un millón de registros para Medellín, se usó el 90 % de los datos para entrenar los modelos; y el 10 % para la evaluación.

Los parámetros del clasificador fueron optimizados haciendo una búsqueda en malla y siguiendo una estrategia de validación cruzada con cinco particiones sobre el conjunto de entrenamiento. El clasificador optimizado en entrenamiento fue evaluado luego en el conjunto de prueba. El modelo de RF tiene tres hiperparámetros que se busca estimar: el número de árboles, la profundidad máxima de los árboles, y el número de características que deben considerarse en cada división. Para ello, se definió una malla de búsqueda con veinte combinaciones de los valores de los hiperparámetros.

El modelo de árboles aleatorios se estimó bajo diferentes tipos de procesamiento de los datos, antes de

la optimización de los modelos, con el fin de buscar el mejor desempeño:

1. Modelos con las clases desbalanceadas.
2. Modelos teniendo en cuenta un balanceo de clases hacia abajo (*undersampling*).
3. Modelos teniendo en cuenta un balanceo de clases hacia arriba (*oversampling*), utilizando la técnica *random oversampling examples* –ROSE (Menardi y Torelli, 2014).

Para la evaluación del desempeño de los modelos de clasificación estimados se usaron seis medidas en el Sisben IV:

1. *Sensibilidad*: mide el porcentaje de personas clasificadas correctamente como pobres.
2. *Especificidad*: mide el porcentaje de personas clasificadas correctamente como no pobres.
3. *Exactitud (accuracy)*: mide el porcentaje de personas pobres y no pobres clasificadas correctamente por el sistema.
4. *Exactitud balanceada*: es un promedio entre la sensibilidad y la especificidad, ponderando con igual importancia ambas medidas.
5. *Área bajo la curva ROC (AUC)*: permite identificar qué tan bien el sistema separa correctamente las muestras clasificadas. El indicador toma valores en el rango (0.5-1.0); y cuanto mayor es el valor, mejor es el modelo para distinguir entre las personas pobres y no pobres.
6. *Pérdida de eficiencia logarítmica*, donde es la probabilidad observada y es la probabilidad real. Esta métrica mide el desempeño del modelo enfocándose en la certeza de la predicción; y cuanto menor es el valor, mejor es la predicción del modelo.

### E. Análisis cualitativo

El análisis cualitativo surgió de la inquietud sobre las personas clasificadas como no pobres por el Sisben IV, que necesitan, sin embargo, ayuda social prioritaria, debido a sus apremiantes condiciones sociales. Para abordar esto, nos acercamos al programa de acompañamiento familiar “Familias Medellín” de la Secretaría

de Inclusión Social de esa ciudad. Este programa utiliza profesionales de las ciencias sociales para acompañar a familias en situación de pobreza y pobreza extrema, con el objetivo de ayudarles a superar esa condición y eliminar la pobreza en la ciudad.

Con el equipo territorial, se identificaron casos donde los hogares estaban clasificados en el Sisben IV en las categorías C y D, es decir, vulnerables y no pobres y no vulnerables, pero con señales de pobreza. Se accedió a información anonimizada de la situación de cuatro familias. Información que se analizó manteniendo el respeto por la confidencialidad. Los casos fueron referenciados por “Familias Medellín”, sin embargo, lo relevante para ser tenidos en cuenta es que estaban clasificados como no pobres en el Sisben IV.

Esta información se usó en el artículo para describir de manera sucinta la vida de las familias y personas que necesitan ayuda social, pero cuya clasificación en el Sisben IV limita sus posibilidades. Los nombres de las personas y las familias mencionados en el artículo corresponden solo a una estrategia de trabajo, asignados arbitrariamente.

## III. Resultados

La promesa del Sisben IV era mejorar la identificación de personas pobres y vulnerables, al integrar la perspectiva de calidad de vida con la de ingresos. Según los resultados de las tablas de contingencia, el Sisben IV identifica como pobre al 37.69% de las personas que no serían pobres según el IPM, mientras que no identifica como pobre al 4.89% de las personas que sí serían pobres según el IPM (cuadro 3). Tomando como referencia el total de registros de la base de datos, que es una muestra del país, se habla de 181 238 personas. No se observan diferencias en los resultados al comparar el Sisben con el IPM modificado con pesos iguales en los indicadores.

En Medellín, al comparar el Sisben con el IPM oficial, en general, se evidencia la misma tendencia que en Colombia. El 14.73% es pobre según el Sisben IV, pero no es considerado pobre según el IPM, mientras que el 3.73% es pobre según el IPM (lo que equivale a 49 806 personas en la ciudad), pero el Sisben no logra identificarlos. Ambos métodos identifican como pobre al 5.82% de las personas.

Estos resultados revelan una baja concordancia entre los métodos para identificar a los pobres, independientemente del IPM evaluado. En Medellín, cuando se consideran pesos iguales para el IPM, los porcentajes cambian aproximadamente en 4 puntos porcentuales en todos los grupos, en comparación con el IPM oficial. No obstante, con pesos iguales, es menor el porcentaje de población pobre por IPM y no pobre por Sisben, es decir, una menor exclusión del Sisben de los pobres por IPM.

### A. Concordancia entre el IPM y el Sisben IV

En la evaluación de la concordancia, cuando el umbral del IPM es bajo, al 10%, se espera que la población tenga un 10% de privaciones en los indicadores ponderados, pero que no sea pobre según el Sisben. En este caso, el nivel de concordancia tendería a ser bajo. Por otro lado, cuando el umbral del IPM es alto, se espera una concordancia igualmente baja, ya que la medida de pobreza se vuelve muy estricta y tiende a excluir a muchos que podrían ser identificados por el Sisben.

De esta manera, es más probable que sean considerados pobres por el Sisben sin que eso implique ser pobres multidimensionalmente, por lo que la concordancia es baja y tiende a cero en umbrales altos del IPM. En los valores medios del umbral de privación o punto de corte, entre el 20 y el 40%, debería alcanzarse la mayor concordancia entre las dos medidas.

Al ajustar la medida con el coeficiente kappa, que permite mejorar el componente de azar de la concordancia estimada (Viera *et al.*, 2005), sugiere que parte de la concordancia básica se debe al azar. En los valores extremos de los umbrales de la pobreza

multidimensional, principalmente en los valores altos, el coeficiente es menor. Mientras, en los cortes entre el 0.2 y el 0.4 del IPM se logra la mayor concordancia de los indicadores (figura 3), tanto para Colombia como para Medellín.

En el comportamiento del índice kappa se evidencia la tendencia esperada con respecto al umbral de pobreza multidimensional. Sin embargo, la concordancia medida tanto para Colombia como para Medellín es baja o insignificante. Los valores más altos del indicador de concordancia están por debajo del 0,36, donde se alcanza el mayor valor con el IPM con pesos iguales, con un punto de corte del IPM igual a 0.2, tanto en Colombia como en Medellín (gráfico 3).

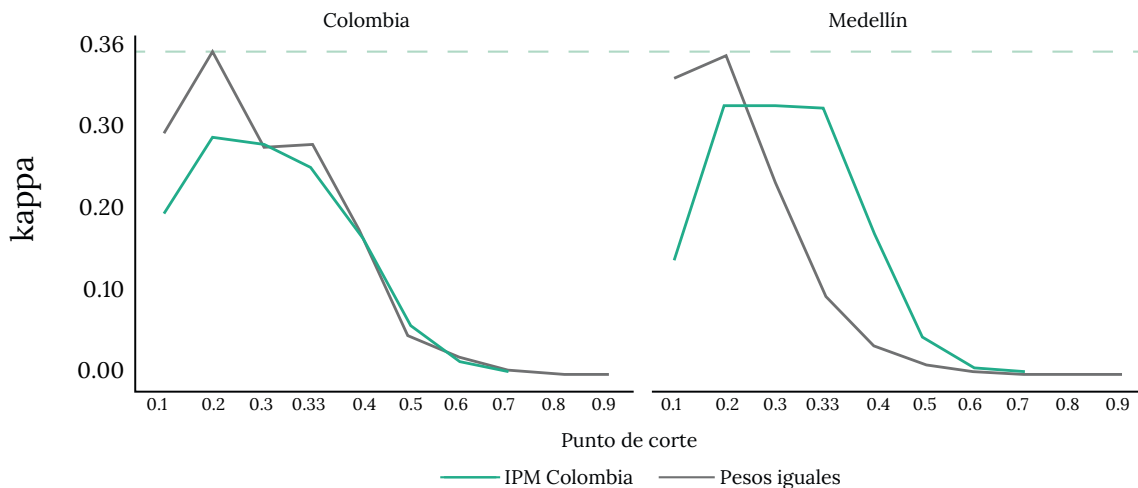
La mayor concordancia entre el Sisben y el IPM de Colombia con los pesos oficiales se encuentra en 0.26, cuando el corte de pobreza multidimensional es igual al 20% de las privaciones ponderadas. Por su parte, el valor más alto de concordancia entre la medida oficial del IPM para Medellín y el Sisben IV llega a 0,3, y se logra con un valor de corte de pobreza multidimensional igual al 30%, cercano al puntaje de corte oficial de la pobreza multidimensional en Colombia (0,33).

El resultado permite evidenciar la sensibilidad de la identificación de los pobres ante cambios en el corte para definir la pobreza multidimensional. A todos los umbrales del IPM, los grupos pobres según ambos métodos son los que presentan menor incidencia de la pobreza; resultado de su baja concordancia. En los diferentes grupos, tanto para Colombia como para Medellín, el punto de corte donde las diferentes medidas evaluadas del IPM identifican el mismo porcentaje de población pobre está en el rango 40-50% de las privaciones ponderadas (figuras 4-5).

**Cuadro 3.** Grupos según la combinación de pobreza por Sisben IV y pobreza multidimensional, en porcentaje (corte del IPM al 0,33 de las privaciones ponderadas)

| Grupos                         | Colombia    |                   | Medellín    |                   |
|--------------------------------|-------------|-------------------|-------------|-------------------|
|                                | IPM oficial | IPM pesos iguales | IPM oficial | IPM pesos iguales |
| Pobres por ambos métodos       | 30.78       | 30.70             | 5.82        | 1,35              |
| Pobres por Sisben y no por IPM | 37.69       | 37.77             | 14.73       | 19.21             |
| Pobres por IPM y no por Sisben | 4.89        | 3.83              | 3.73        | 0.47              |
| No pobres por ambos métodos    | 26.64       | 27.69             | 75.72       | 78.97             |

Fuente: elaboración propia.



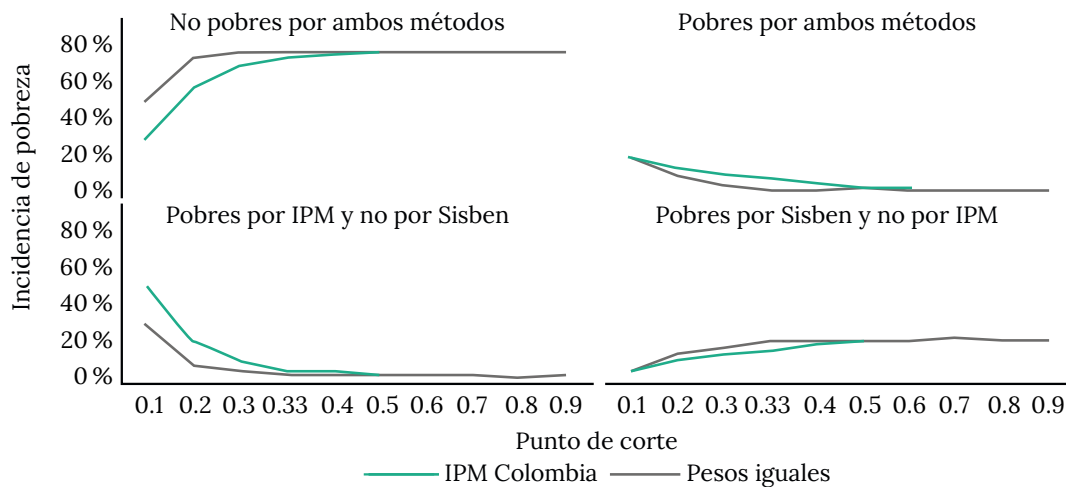
**Figura 3.** Concordancia kappa entre el IPM y Sisben IV, cambiando el umbral de pobreza multidimensional. IPM oficial y con pesos iguales para Colombia y Medellín

Fuente: elaboración propia, a partir de datos del Sisben.

En el grupo particular de pobres por IPM y no pobres por Sisben, para Medellín, se observa que a partir del corte del 50% de las privaciones ponderadas, el porcentaje de población es cercano a cero, en el IPM oficial y el de pesos iguales. En Colombia, en el mismo grupo y para los dos índices de pobreza evaluados, el valor donde el grupo se acerca a cero se encuentra en el corte del 40% de las privaciones ponderadas. Es en esos puntos de corte donde el Sisben lograría absorber a los pobres multidimensionales. Conviene

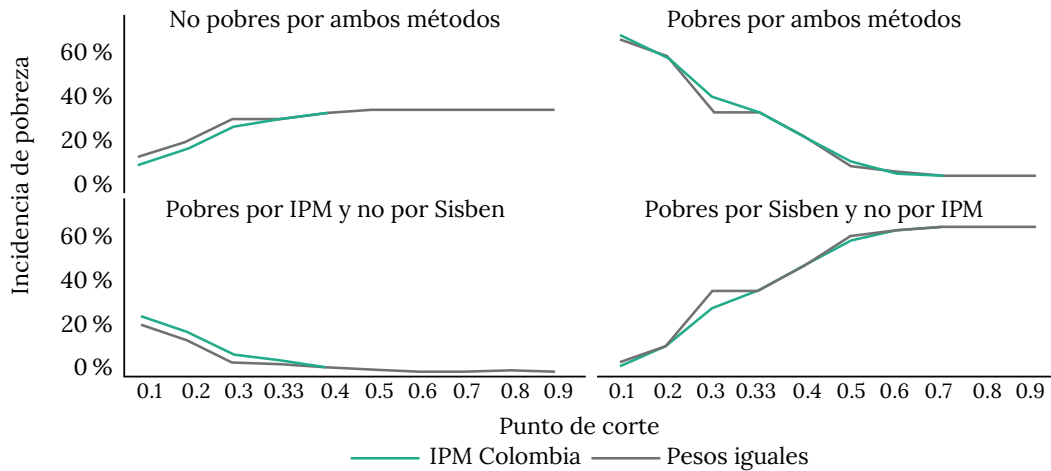
recordar que el punto de corte de la medición oficial de Colombia es del 33%.

Además, en las figuras 4-5 se observa que, en los diferentes grupos, tanto para Colombia como Medellín, el ordenamiento de las diferentes medidas, de acuerdo con la incidencia de la pobreza, tiende a mantenerse a medida que varía el puntaje de corte del IPM. Las medidas de IPM analizadas son robustas, ya que se mantiene el ordenamiento de las mediciones, cuando cambia el punto de corte.



**Figura 4.** Grupos de la tabla de contingencia y variaciones de umbral de corte del IPM en Medellín, la incidencia de la pobreza describe el porcentaje de población

Fuente: elaboración propia, a partir de datos del Sisben.



**Figura 5.** Grupos de la tabla de contingencia y variaciones de umbral de corte del IPM en Colombia, la incidencia de la pobreza describe el porcentaje de población

Fuente: elaboración propia, a partir de datos del Sisben.

### B. Estimación de pobreza por Sisben IV

Los resultados de los mejores modelos estimados se presentan en la tabla 4. El modelo con el mejor rendimiento balanceado se obtuvo mediante el proceso de muestra reducida, con un acierto balanceado de 0.78, sensibilidad de 0.81, especificidad de 0.76, acierto de 0.77, AUC de 0.86 y  $Mn \log loss$  igual a 0.47. El modelo estimado sin balancear las clases presenta los resultados esperados: un alto nivel de acierto, debido a la mayor capacidad de predecir la clase no pobre. Esto era esperado porque, durante el entrenamiento, el modelo reveló un mayor número de observaciones clasificadas como no pobres, en comparación con las clasificadas como pobres.

Entre los modelos estimados con muestra reducida y con sobremuestra, los valores de las medidas de desempeño son cercanos. Sin embargo, se eligió el modelo con el mejor acierto balanceado, en la medida en que se buscaba lograr un equilibrio en la capacidad de predecir tanto la clase pobre como la no pobre. Además, con la presente estimación, se intentó emular la clasificación realizada por el Sisben IV, resultado que se evidencia mejor con la muestra reducida, donde la sensibilidad y la especificidad están más equilibradas.

A partir del modelo con muestra reducida, se realizó el análisis de las variables de importancia, el cual refleja la capacidad de predicción de las variables utilizadas en el modelo. Es necesario advertir que los resultados

**Tabla 4.** Resultados del modelo de árboles aleatorios con diferentes tratamientos de balanceo de las clases pobre y no pobre

|   | Modelo estimado  | Sensibilidad | Especificidad | Acierto | Acierto balanceado | ROC AUC     | Perdida logarítmica de la eficiencia |
|---|------------------|--------------|---------------|---------|--------------------|-------------|--------------------------------------|
| 1 | Sin balancear    | 0.54         | 0.91          | 0.81    | 0.73               | 0.86        | 0.41                                 |
| 2 | Muestra reducida | <b>0.81</b>  | <b>0.76</b>   | 0.77    | <b>0.78</b>        | <b>0.86</b> | 0.47                                 |
| 3 | Sobre muestra    | 0.73         | 81            | 0.79    | 0.77               | 0.86        | 0.44                                 |

Fuente: elaboración propia.

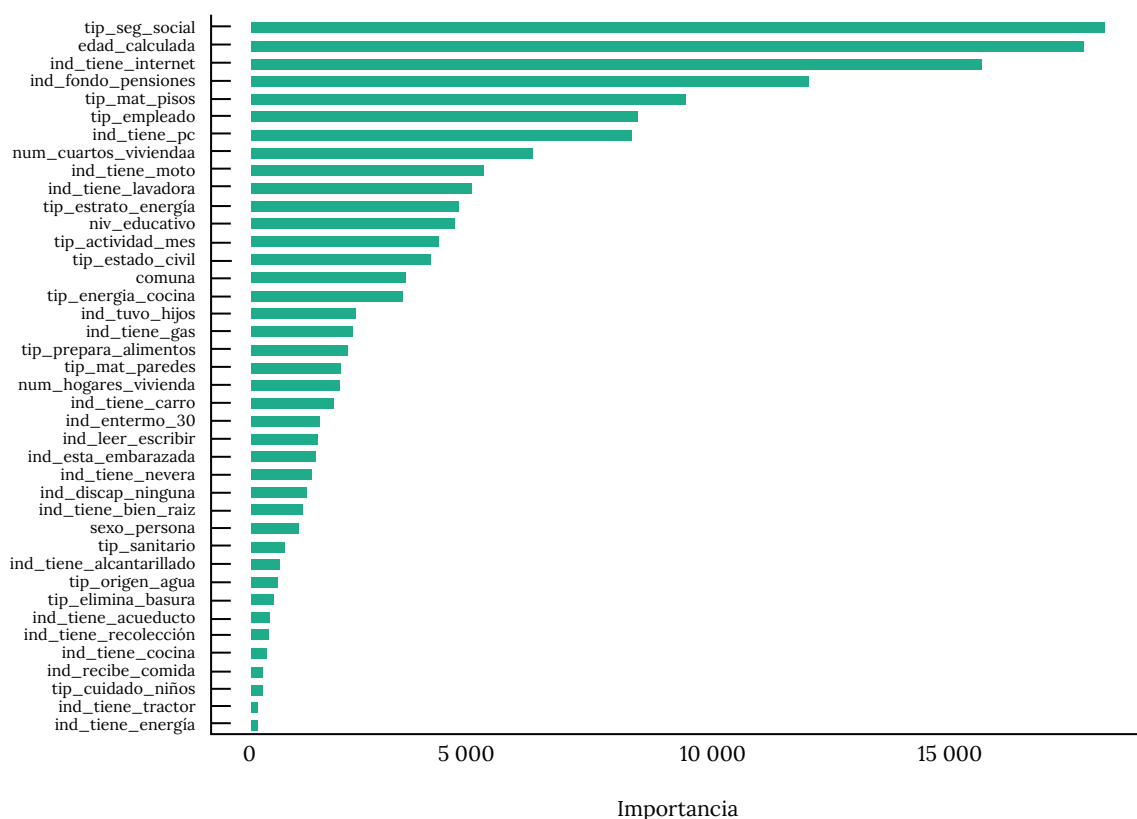
son particulares, debido a la presencia de las diferentes variables que acompañaron la estimación. En la figura 10 se presentan las variables en orden de importancia, cuyo valor más alto significa mayor capacidad de predicción de la variable.

El gráfico muestra el total de variables utilizadas en el modelo de clasificación (40). Las cinco más importantes para la estimación son el tipo de seguridad social, la edad de la persona, la tenencia de internet en el hogar, si la persona cotiza en un fondo de pensiones y el material de los pisos de la vivienda. En las diez primeras variables de importancia, destacan los indicadores de tenencia de activos y los asociados al mercado laboral. Las variables con menor capacidad predictiva son: la tenencia de energía, la tenencia de tractor, el cuidado de los niños (indaga por dónde o con quién permanecen el niño o la niña más tiempo entre semana), si el niño o la niña reciben comida en

el lugar donde pasan la mayor parte del tiempo; y la tenencia de cocina.

Al evaluar las métricas de desempeño del modelo para el grupo de pobres y no pobres por IMP, se observan resultados similares (cuadro 5). Esto significa que el modelo estimado no está sesgado por la variable de pobreza multidimensional. Ser pobre o no pobre multidimensionalmente no influye en que el modelo genere mejor o peor predicción. El modelo es consistente para identificar ambas clases, independientemente de la característica de ser pobre o no pobre de forma multidimensional.

Ahora bien, al analizar el modelo de acuerdo con los grupos de pobres o no pobres por Sisben IV (cuadro 6), se predice mayor proporción de población pobre según el Sisben IV en comparación con los valores reales (balanceados). La proporción de casos predichos es mayor en



**Figura 6.** Ranking de las 40 variables de acuerdo con su capacidad de predicción de los grupos pobre (A+B) y no pobres (C+D) del Sisben IV

*Nota.* Los valores altos indican mayor importancia.

*Fuente:* elaboración propia a partir de la base de datos del Sisben IV para Medellín.

**Cuadro 5.** Métricas del desempeño en grupos de pobre y no pobre por IPM

| Pobre por IPM | Acierto | Sensibilidad | Especificidad | Acierto balanceado |
|---------------|---------|--------------|---------------|--------------------|
| No            | 0.77    | 0.76         | 0.81          | 0.78               |
| Sí            | 0.77    | 0.76         | 0.82          | 0.79               |

Fuente: elaboración propia.

**Cuadro 6.** Comparación entre los grupos de pobreza por Sisben y por IPM de acuerdo con los valores reales y los predichos por el modelo estimado

| Grupos                         | Grupo predicho (%) | Grupo real (%) |
|--------------------------------|--------------------|----------------|
| Pobres por ambos métodos       | 7.17               | 4.87           |
| Pobres por Sisben y no por IPM | 32.55              | 22.31          |
| Pobres por IPM y no por Sisben | 10.79              | 13.09          |
| No pobres por ambos métodos    | 49.49              | 59.73          |

*Nota.* El IPM es la medida oficial calculada para Medellín. Los valores son diferentes a los reportados en el cuadro 3, porque el modelo se estimó con las clases del Sisben balanceadas.

Fuente: elaboración propia.

los grupos de pobres por ambos métodos y pobres por Sisben, pero no por el IPM: con 7.17% y 32.55%, respectivamente, en comparación con 4.87% y 22.31% de los valores reales. El modelo tiende a clasificar mayor porcentaje de población como pobre, según el Sisben (32.55%, en relación con el real de 22.31%). El algoritmo asigna menor porcentaje de los pobres por IPM y no por Sisben, así como a los no pobres por ambos métodos. Cabe resaltar que, con la estimación, no es posible determinar en qué grupo de pobreza queda la predicción, según la clasificación del Sisben.

### C. Realidades de ser no pobre según el Sisben IV

Una de las principales discusiones sobre los algoritmos que evalúan la calidad de vida de las personas es su capacidad para captar las realidades con precisión. El instrumento del Sisben IV es un algoritmo que determina el tipo de ayuda social que una persona o familia merece recibir, según su capacidad para predecir la pobreza y la vulnerabilidad. Sin embargo, es la realidad de la vida de las personas la que permite centrar los juicios y ajustar los algoritmos. Así pues, en este apartado, se describen las realidades de cuatro familias

que solicitan ayuda del Estado y que se encuentran en situaciones que no consideradas pobres y, en algunos casos, no vulnerables, según la clasificación del Sisben IV.

#### 1. Familia Naranjales

Andrés Naranjales se ve forzado a acudir a la defensoría del pueblo para solicitar apoyo, debido a su condición. Tiene 36 años, es soltero y cuenta con educación técnica en el área comercial. Perteneció a la comunidad LGBTIQ+; diagnosticado con VIH desde hace cuatro años. Además, se encuentra desempleado.

Manifiesta que solicitó apoyo en forma de ayuda social, debido a su falta de empleo y a los constantes ataques psicológicos y xenófobos que recibe asociados al género. También ha perdido su red de apoyo familiar. Sus problemas de salud le producen deterioro físico y psicológico, requiere de medicación constante y atención médica, de la cual manifiesta estar privado, debido a los frecuentes retrasos y a la dificultad para acceder a citas médicas. Andrés oculta el diagnóstico de la enfermedad a su familia, excepto a su madre; y pide prudencia con la información, para evitar que se

enteren, ya que la violencia psicológica podría intensificarse. Por solicitud expresa de Andrés, el informe solo se refiere a él y no a su grupo familiar.

El hogar está conformado por los padres y hermanos de Andrés. Sin embargo, él manifiesta que, principalmente los hermanos no le ofrecen apoyo y se convierten en agresores psicológicos debido a su orientación sexual y su imposibilidad para contribuir económicamente. El hogar enfrenta limitaciones para generar ingresos suficientes para la alimentación y otras necesidades básicas. La vivienda pertenece a los padres, pero Andrés debe contribuir para pagar los servicios y cubrir los gastos relacionados con su enfermedad.

Además de no contar con fuentes de ingresos, ha iniciado un emprendimiento de venta de postres, pero manifiesta tener pocas ventas y no poder cubrir los costos fijos. El mercado laboral lo excluye debido a su condición de salud y al diagnóstico de VIH. Andrés manifiesta, pues, tener pensamientos suicidas que se intensifican debido a las dificultades que enfrenta.

Andrés se encuentra en el Sisben IV, clasificado en el grupo C y en el subgrupo 14 de los 18 disponibles. El grupo C lo clasifica como vulnerable, pero el subgrupo proporciona información sobre la profundidad de la vulnerabilidad. El resultado de 14 implica que Andrés tiene una mejor calidad de vida y mayor capacidad para generar ingresos, en comparación con alguien ubicado en el mismo grupo, pero en un subgrupo inferior. Andrés se encuentra a dos subgrupos de ser considerado ni pobre ni vulnerable según el Estado colombiano. En este caso, no es una persona a la que se priorice para la atención social, ya que por detrás del grupo C se encuentran los grupos B y A, que suman 12 subgrupos en total.

## 2. Familia López

Los López son una familia de cuatro personas. Yaneth tiene 36 años y es madre cabeza de hogar. Su esposo tiene 39 años. Tienen una hija de 8 años y viven con la abuela, madre de Yaneth, que tiene 66 años. Yaneth solo cuenta con educación primaria, y trabaja en la temporada decembrina vendiendo calzado en el centro de Medellín. Es un trabajo temporal por días. El esposo completó la educación primaria y está registrado como víctima de desplazamiento forzado. Trabaja en varios oficios de construcción y gana un

salario mínimo, que la familia utiliza para cubrir todos los gastos del hogar.

Yaneth toma medicamentos para el colesterol y su madre sufre de hipertensión. Su hija debe usar frecuentemente un inhalador para tratar problemas de rinitis e infecciones respiratorias. La familia tiene dificultades para alimentarse adecuadamente; y solo puede acceder a alimentos de escaso valor nutricional.

Por otro lado, la casa donde viven es propia, pero la están pagando a través de un crédito, cuya cuota, con frecuencia, les resulta difícil de pagar. La casa tiene problemas de humedad, malos olores y filtraciones de agua en las paredes. Obtienen electricidad a través de un sistema prepago, pero a veces no pueden recargar, porque no tienen suficiente dinero. Los ingresos del hogar son insuficientes para cubrir las necesidades básicas. Sin embargo, su clasificación en el Sisben IV es grupo C, subgrupo 2. Esto significa que, según el Estado, la familia Gutiérrez no se encuentra en pobreza extrema ni en pobreza moderada. El instrumento los clasifica como vulnerables.

## 3. Familia Gómez

La familia está conformada por Juan Camilo, de sesenta años, que vive en una zona rural de Medellín. Es separado y vive solo. Juan Camilo es un adulto mayor que trabaja de manera informal como zapatero y reciclador de residuos sólidos. Su nivel educativo alcanza la educación básica primaria.

En las labores que realiza, obtiene ingresos diarios de 10 000–40 000 pesos colombianos, además de los ingresos esporádicos por la venta de los residuos que recupera. Manifiesta que sus gastos mensuales básicos son de 80 000 pesos en alimentación, 23 000 pesos en energía prepago, 33 000 pesos en agua y alcantari-lado y 17 500 pesos en alimento para sus animales de compañía (gatos y perros). Aunque vive solo, cuenta con el apoyo de una hermana que le proporciona alimentos preparados.

El psicólogo que visitó su vivienda evidenció un posible trastorno de acumulación compulsiva. Juan Camilo enfrenta denuncias de los vecinos por olores e invasión de la zona pública con basura. No puede desplazarse adecuadamente en su vivienda debido a la cantidad de residuos acumulados. Se observaron problemas de



roedores, moluscos e insectos, además de tener gatos y perros viviendo con él.

Debido a su situación, Juan Camilo manifiesta sufrir de angustia con frecuencia, situación que se agrava al vivir solo. Aunque tiene acceso a la atención médica subsidiada, no cuenta con la vacunación para contraer la gripa Covid-19. Solo puede comer dos veces al día y su dieta se basa en carbohidratos y productos altos en colesterol.

Juan Camilo no cumple con los requisitos para solicitar el subsidio de la estrategia Colombia Mayor, que se enfoca en mujeres mayores de 54 años y en los grupos del Sisben en las categorías A1 a C1. La clasificación que obtuvo en el Sisben IV es D9, lo que significa que se encuentra en el grupo D, que lo etiqueta como no pobre y no vulnerable. Además, se encuentra en el subgrupo 9, clasificación que le impide ser priorizado por los diferentes programas sociales de Medellín y Colombia.

#### 4. Familia Álvarez

La familia Álvarez es un hogar unipersonal. Doña Beatriz es una adulta mayor soltera, de 67 años, que reside en la zona urbana de Medellín. Solo alcanzó a cursar educación primaria. Su caso se conoce a través de un inspector de policía, quien llevó a cabo un proceso administrativo en su contra por incumplimiento de la Ley 1801 de 2016 sobre el Código Nacional de Seguridad y Convivencia Ciudadana, debido a que no cumple con normas de seguridad y sanidad. Este proceso se debe a que tiene residuos acumulados en el exterior de la vivienda, lo cual genera problemas de salud pública.

Beatriz trabaja como recicladora informal, durante 7 a 8 horas diarias. No forma parte de ninguna cooperativa o asociación y ejerce su función de manera independiente. Lleva 38 años viviendo en el sector y está sola desde 2012, cuando grupos al margen de la ley asesinaron a su hijo; también falleció su madre en ese mismo año. Manifiesta que, con frecuencia, olvida la información que le brindan las personas que la visitan.

El psicólogo que realizó la visita observa un comportamiento de acumulación compulsiva por parte de Beatriz. Tanto el interior como el exterior de su vivienda están llenos de plásticos, cartón, vidrio, textil,

calzado, chatarra, madera y otros, apilados en diferentes partes. Beatriz tiene dificultades para desplazarse por la vivienda y, aunque cuenta con un lugar para dormir, las condiciones de humedad y malos olores son intensas.

En 2021, Beatriz sufrió un incendio en su hogar debido a un accidente con el fogón, lo que causó la destrucción de sus electrodomésticos y enseres, incluyendo la red eléctrica. Desde ese incidente, no cuenta con los enseres básicos ni con un lugar donde preparar sus alimentos; de manera que compra su comida en la calle.

Tiene acceso a agua en la vivienda, que paga mediante sistema prepago. La infraestructura en general se encuentra en mal estado. Su red de apoyo es muy débil. Menciona que tiene dos nietas, pero no se comunica frecuentemente con ellas. Y los vecinos le ayudan esporádicamente con alimentación. Doña Beatriz solo puede comer dos veces al día y ha visto disminuido su acceso a la alimentación, debido al aumento de los precios. En esas condiciones, su dieta no cumple con los requisitos mínimos de nutrición, por ser alta en colesterol y carbohidratos.

Se encuentra activa en el sistema de salud subsidiado, aunque menciona que hace varios años no acude a consulta. Afirma haber recibido una dosis de la vacuna contra el Covid-19. Es una adulta mayor que ha atravesado diversas situaciones que han afectado negativamente su bienestar. No está clasificada como posible beneficiaria de ayuda social estatal, ya que su grupo en el Sisben IV es C y su subgrupo es el 16. Es decir, se considera vulnerable pero no pobre, y se encuentra a dos subgrupos de pasar al grupo D, que implicaría no ser pobre ni vulnerable.

## IV. Discusión

Los resultados de la concordancia entre el Sisben IV y la pobreza multidimensional, tanto a nivel nacional como de ciudad (Medellín), revelan una baja o insignificante identificación conjunta de la población pobre. Ambas medidas muestran una capacidad limitada para identificar de manera simultánea a las personas en situación de pobreza. Es particularmente notable la proporción significativa de población que, a pesar de no ser considerada pobre según el IPM, es clasificada como pobre por el Sisben.

Es necesario revisar el argumento presentado en el documento de política social del Gobierno nacional (DNP, 2016) que relaciona la medida del Sisben IV con el IPM, a fin de evitar la exclusión de la población pobre identificada por el IPM, pero no por el Sisben. Las discrepancias entre ambas medidas pueden estar vinculadas a la necesidad del Sisben de dar respuesta a las dos formas oficiales de medición de la pobreza: el IPM y el enfoque basado en los ingresos. Por tanto, una hipótesis es que las personas consideradas pobres por el Sisben, pero no por el IPM, podrían estar experimentando pobreza monetaria. Otra razón podría estar relacionada con la naturaleza del algoritmo del Sisben, el cual identifica a los pobres de manera independiente de las mediciones oficiales de pobreza en el país.

El valor de concordancia más alto obtenido entre el Sisben IV y el IPM, en las diferentes mediciones, resulta menor que el reportado para Colombia y otros países de América Latina en estudios que comparan la línea de pobreza monetaria y el indicador de carencias sociales (Villatoro y Santos, 2019). En las regiones evaluadas en el estudio citado, la concordancia es a lo sumo moderada. La baja concordancia reportada entre el Sisben y el IPM advierte de que ambas mediciones capturan aspectos diferentes del fenómeno de la pobreza. Al considerar el resultado del Sisben IV como una medida exógena, con validez para la medición de la pobreza, se reconoce que el proceso metodológico tiene decisiones normativas que afectan la concordancia con las medidas de pobreza multidimensional.

Por tanto, es esperado que las dos mediciones de pobreza no concuerden al 100 %. Sin embargo, es responsabilidad de la política pública de Colombia atender a las personas que quedan excluidas de la ayuda social del Estado, si son pobres multidimensionales. En porcentaje, representan el 4.89 % para Colombia y el 3.73 % para Medellín, ambos según la comparación del IPM oficial. Con los resultados, ni el Sisben ni el IPM parecen ser suficientes por sí solos para evidenciar la pobreza en Colombia y, más concretamente, en Medellín.

De los resultados, en particular, se deriva la importancia de contar con mediciones que se usen de forma complementaria para la identificación de la población pobre. Dado que el método del Sisben IV estima la capacidad de generar ingresos, los resultados

presentados aquí advierten la misma conclusión que otros estudios, donde se analizan la pobreza monetaria y la multidimensional, a saber, las personas de bajos ingresos resultan no ser las mismas identificadas como multidimensionalmente pobres (Alkire y Seth, 2008; Alkire y Shen, 2017; Roelen, 2017; Tran *et al.*, 2015).

Al integrar varias mediciones de pobreza se enriquece la discusión sobre la pregunta *¿quién es pobre?*, pero deja un reto a la política pública para definir arreglos institucionales que no dejen a ninguna persona pobre sin la posibilidad de recibir apoyo de la política social.

En el caso particular de Colombia, las medidas de pobreza que deben integrarse son tres. El proceso de evaluar la pobreza con el indicador multidimensional y el de ingresos hace necesario que la medida de focalización de la pobreza sea una agregación de ambas y cumpla su función de ser una medida superior; e incluir las diferentes medidas de pobreza.

Por consiguiente, los resultados sugieren que la clasificación de las personas en los grupos A y B del Sisben IV debe complementarse con la medición de la pobreza multidimensional, con el fin de no dejar a nadie atrás, desde la premisa del primer ODS: “Poner fin a la pobreza en todas sus formas en todo el mundo” (United Nations, 2015).

Aunque el artículo no se enfocó en la discusión de los pesos de los indicadores para definir la estructura de la evaluación del IPM, se evidencian algunos aspectos. El cambio de la estructura de pesos de los indicadores del IPM, no genera cambios significativos en los resultados para los cuatro grupos evaluados: pobres por ambos métodos, pobres por Sisben y no por IPM, pobres por IPM y no por Sisben, y no pobres por ambos métodos. En Medellín, se observó un cambio proporcional grande al considerar pesos iguales para el IPM, particularmente, en los grupos de pobres por ambos métodos y pobres por IPM y no por Sisben. Al parecer, en Colombia, se obtienen resultados consistentes a nivel agregado y, en Medellín, se altera la composición de los porcentajes de los grupos analizados.

El resultado de la estimación del modelo muestra la capacidad del algoritmo RF para clasificar la población pobre y no pobre por el Sisben IV. Las variables más importantes tienen coherencia con la esencia del Sisben IV, en cuanto a su declaración de evaluar la

capacidad de generar ingresos. Las variables con mayor importancia para la estimación fueron las asociadas con el mercado laboral y las de tenencia de activos.

Con el resultado se comprueba el énfasis del algoritmo del Sisben en aspectos complementarios a los que se observan con la medición del IPM. Por ejemplo, se observa distancia en los aspectos educativos, en los servicios con los que cuenta la vivienda, en variables de salud y cuidado de los niños y en las variables de activos. Esta última, no se contempla por el IMP oficial, pero es importante para la clasificación del Sisben.

En este sentido, mantener la focalización de la ayuda social bajo la perspectiva de la capacidad de generar ingresos es una postura que requiere ampliarse. La evidencia que refuerza la premisa se ilustra con los resultados del análisis cualitativo de las familias, clasificadas como no pobres por el Sisben IV, pero a la luz de un escrutinio social se podría llegar con facilidad a la conclusión de su condición de pobreza.

Las familias que hicieron parte del estudio cualitativo presentan señales claras de pobreza, pero el algoritmo del Sisben IV no logra identificarlos como pobres, de modo que los excluye de la ayuda social estatal. El reto es llevar a cero la exclusión de los pobres de la ayuda social, vigilando que los llamados “colados” en el sistema sean los menos posibles. En este punto, la estrategia declarada debe preferir identificar bien a los pobres y, en segunda instancia, evitar a los colados en el sistema.

En el artículo, se reporta un caso de exclusión que va más allá de los expuestos por la literatura de algoritmos de inteligencia artificial y aprendizaje de máquinas (Corbett y Goel, 2018). La discriminación a la que nos referimos se hace sobre la base de un constructo complejo la pobreza, el cual se constituye en un acuerdo moral y normativo que puede o no incluir las características de identificación directa como el género/sexo, la raza, la etnia, el lenguaje, la religión o creencias, la edad, la opinión política, la discapacidad o el lugar de residencia, comunes en la evaluación de la discriminación por algoritmos de inteligencia artificial o aprendizaje de máquinas (Barocas *et al.*, 2017; Celis *et al.*, 2019; Chen *et al.*, 2021; Corbett y Goel, 2018; Eubanks, 2018; Keddell, 2019; Lepri *et al.*, 2017).

Por consiguiente, la prueba de la concordancia, sumada a la evaluación cualitativa, permite reconocer la necesidad de mejorar la clasificación automática que realiza el Sisben IV en la actualidad. Precisamente, el análisis cualitativo es una forma para mejorar los resultados de la clasificación por algoritmos (Raji *et al.*, 2020).

El artículo también abre la discusión y la oportunidad para implementar un proceso de auditoría interna y gobernanza algorítmica en el DNP de Colombia y su contraparte en Medellín, para anticiparse a las consecuencias negativas antes de que ocurran, como se viene discutiendo en términos de la implementación de algoritmos de aprendizaje de máquinas y la inteligencia artificial (Chouldechova y Roth, 2020; Corbett y Goel, 2018; Raji *et al.*, 2020).

Lo anterior implica una declaración ética de los resultados y efectos que deben ser rechazados por los posibles efectos sociales negativos sobre las personas que se busca beneficiar. Aunado al reconocimiento de la importancia de validar el significado de la pobreza y la postura de equidad para recibir la ayuda social del Estado, desde un acuerdo social que se fundamente en deliberaciones razonadas de la sociedad (Sen, 2018).

Son varias las limitaciones del presente estudio, que sirven, sin embargo, para seguir avanzando en las áreas de investigación de los indicadores de identificación de la pobreza. Primero, la base de datos del Sisben IV tiene restricciones:

1. No tienen las variables para calcular el índice de pobreza monetaria. Eso limita el alcance de la comparación con el IPM
2. Con las variables del Sisben, se reproduce una medición aproximada del IPM oficial.
3. Para el caso de Colombia, las bases de datos de familias, hogares y personas no tienen un indicador que permita unirlos.
4. Para Medellín, la base no cuenta con los registros de la población que habita las zonas rurales.

Segundo, en el tratamiento de los datos, redujimos los grupos del Sisben IV a una variable dicotómica (pobres y no pobres), lo que evita analizar las brechas entre los pobres y los indicadores de desigualdad de la pobreza. Tercero, al no contar con las variables que explican la

clasificación de los grupos del Sisben IV, las comparaciones entre los grupos dados por la relación entre las dos medidas son limitadas.

En relación con lo anterior, también puede listarse como limitación final que los modelos de aprendizaje estadístico son sensibles a los tipos de dato y, por consiguiente, otras estimaciones pueden dar diferentes ordenaciones según la importancia de las variables. Sin embargo, de acuerdo con el proceso de estimación del presente artículo, se espera que las variables listadas como importantes para predecir los grupos de pobres y no pobres se encuentren entre las principales.

## V. Conclusiones

La identificación de los pobres, de acuerdo con el Sisben IV y el índice de pobreza multidimensional, tiene una concordancia insignificante. Los mayores valores del indicador de concordancia kappa están por debajo de 0.36, con IPM con pesos iguales y un punto de corte del IPM igual a 0.2, que alcanza el valor más alto. Estas medidas evidencian la pobreza desde diferentes perspectivas de evaluación.

La clasificación de la pobreza del Sisben IV parece ser superior al IPM, ya que incluye aproximadamente un 38% de pobres que no son pobres según el IPM. Sin embargo, deja de identificar como pobres a alrededor del 4 al 5% de las personas pobres multidimensionales. En Medellín, el Sisben incluye al 14.73% que no es pobre según el IPM, pero deja de identificar al 3.73% que es pobre según el IPM. El grupo de personas pobres según el IPM y no pobres según el Sisben comienza a ser cero cuando el corte de las privaciones ponderadas del IPM es cercano al 40%, tanto para Colombia como para Medellín. Cabe recordar que el corte oficial para la pobreza multidimensional es del 33%.

Los casos reportados de las cuatro familias clasificadas como no pobres por el Sisben IV evidencian indicios de formas de pobreza multidimensional. Estas familias son ejemplos de la cotidianidad que enfrentan las familias en la ciudad de Medellín y que son excluidas por el Sisben. La descripción de algunos aspectos de la vida de las personas permite dar cuenta de las diferencias que cada familia tiene en la valoración de la vida, sus restricciones y logros.

La falta de ingresos, la soledad o las posibles enfermedades mentales se capturan mediante la descripción cualitativa. Con el análisis se caracteriza la multidimensionalidad de la pobreza, la interdependencia de las diferentes dimensiones, al tiempo que se observan las diferencias entre los casos analizados. Estas familias consideran que merecen la ayuda social, debido a su condición, y son casos ilustrativos donde el algoritmo del Sisben requiere flexibilidad como instrumento de focalización de la política pública social.

El modelo de RF estimado para clasificar los grupos de pobres y no pobres por el Sisben IV, utilizando datos de Medellín, alcanza un indicador de exactitud balanceada del 78%. El mejor modelo se obtiene con una muestra balanceada, reduciendo la muestra de la clase con mayor representación en la base de datos, es decir, la clase no pobre. El modelo tiene la capacidad predictiva del 81% de los pobres y del 76% de los no pobres.

El modelo se estimó utilizando cuarenta variables asociadas con temas demográficos, vivienda, salud, mercado laboral y tenencia de activos. Las variables de mayor importancia para la predicción fueron principalmente del tema de mercado laboral y tenencia de activos. Las diez principales variables de importancia, según el modelo estimado, son el tipo de seguridad social, la edad de la persona, la tenencia de internet, si la persona cotiza en un fondo de pensiones, el material de los pisos de la vivienda, el tipo de empleo de la persona, la tenencia de computadora, el número de cuartos en la vivienda, la tenencia de moto y la tenencia de lavadora.

Así también, las variables con menor capacidad predictiva son la tenencia de energía, la tenencia de tractor, el cuidado de los niños (indaga sobre dónde o con quién permanece el niño o niña más tiempo entre semana), si el niño o niña recibe comida en el lugar donde pasa la mayor parte del tiempo y, la que tiene menos capacidad para predecir el grupo de pobres o no pobres según el Sisben, es la tenencia de cocina.

En Colombia, se recomienda utilizar la pobreza multidimensional como complemento para la focalización de la política pública social, actualmente basada únicamente en el Sisben IV. Es decir, combinar la información de una medida superior, como el Sisben, pero que no excluya a las personas pobres según las medidas oficiales. Para eliminar la exclusión de la población que

requiere ayuda social, se recomienda a los responsables de la política pública complementar la medición del Sisben IV y el IPM con otras medidas que ayuden a identificar a la población pobre y vulnerable en sus territorios, como bien se plantea en el Conpes 3877.

Ello implica promover ejercicios participativos para lograr una mejor comprensión y contextualización de la pobreza en su medición (Atkinson, 2019). Asimismo, se recomienda complementar los algoritmos de clasificación con perspectivas de análisis que incluyan la lectura del territorio y las diversas manifestaciones de la pobreza. En este sentido, los indicadores para identificar la pobreza deben evitar la objetividad obtusa que los lleva a considerar una variable a lo socioespacial (Sen, 1992).

## Financiamiento y agradecimientos

La realización de la investigación no tuvo fuentes de financiamiento. El artículo es producto de la investigación doctoral del primer autor, intitulada “Identificación de dimensiones y estimación de la pobreza multidimensional utilizando modelos de aprendizaje estadístico y procesamiento de lenguaje natural”, realizada para optar al título de Doctor en Ingeniería Electrónica y de Computación, por la Universidad de Antioquia. Se agradece especialmente a los evaluadores anónimos de la revista por sus valiosos comentarios y sugerencias que contribuyeron significativamente a la mejora del artículo.

## Referencias

- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8), 476-487.
- Alkire, S., Kanagaratnam, U., & Suppa, N. (2020). *The global multidimensional poverty index (MPI) 2020*. <https://www.ophi.org.uk/wp-content/uploads/OPHIMPIMN492020.pdf>
- Alkire, S., Roche, J. M., Ballon, P., Foster, J., Santos, M. E., & Seth, S. (2015). *Multidimensional poverty measurement and analysis*. Oxford University Press.
- Alkire, S., & Seth, S. (2008). *Measuring multidimensional poverty in India: A new proposal* (Working Paper 15). Oxford Poverty and Human Development Initiative.
- Alkire, S., & Shen, Y. (2017). Exploring multidimensional poverty in China: 2010 to 2014. En *Research on economic inequality*. Emerald Publishing Limited.
- Angulo, R., Díaz, Y., & Pardo, R. R. (2016). The Colombian multidimensional poverty index. Measuring poverty in a public policy context. *Social Indicators Research*, 127, 1-38.
- Atkinson, A. B. (1987). On the measurement of poverty. *Econometrica: Journal of the Econometric Society*, 55(4), 749-764.
- Atkinson, A. B. (2019). *Measuring poverty around the world*. Princeton University Press.
- Banco Mundial —BM (2022). *Poverty and shared prosperity 2022: Correcting course*. The World Bank.
- Barocas, S., Bradley, E., Honavar, V., & Provost, F. (2017). Big data, data science, and civil rights. *arXiv preprint arXiv:1706.03102*.
- Boltvinik, J. (2013). Medición multidimensional de la pobreza. América Latina de precursora a rezagada. *Revista Sociedad y Equidad*, 5, 4-29.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Celis, L. E., Huang, L., Keswani, V., & Vishnoi, N. K. (2019). Classification with fairness constraints. A meta-algorithm with provable guarantees. En *Proceedings of the conference on fairness, accountability, and transparency* (pp. 319-328). Association for Computing Machinery.
- Chen, R. J., Chen, T. Y., Lipkova, J., Wang, J. J., Williamson, D. F. K., Lu, M. Y., & Mahmood, F. (2021). Algorithm fairness in ai for medicine and healthcare. *ArXiv, abs/2110.00603*.
- Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5), 82-89.
- Conover, W. J. (1999). *Practical nonparametric statistics* (vol. 350). John Wiley & sons.
- Corbett, S., & Goel, S. (2018). The measure and mismeasure of fairness. A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*.
- Decancq, K., & Lugo, M. (2013, 01). Weights in multidimensional indices of wellbeing. An overview. *Econometric Reviews*, 32(1), 7-34. <http://dx.doi.org/10.1080/07474938.2012.690641>
- Departamento Nacional de Planeación —DNP. (2012). *Metodologías oficiales y arreglos institucionales para la medición de la pobreza en Colombia*. Consejo Nacional de Política Económica y Social. <https://colaboracion.dnp.gov.co/cdt/conpes/social/150.pdf>
- Departamento Nacional de Planeación —DNP. (2016, 5 diciembre). *Declaración de importancia estratégica*

- del sistema de identificación de potenciales beneficiarios (Sisben IV). Consejo Nacional de Política Económica y Social. <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3877.pdf>
21. Eubanks, V. (2018). *Automating inequality. How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
  22. Greco, G. (2018). Setting the weights: The women's capabilities index for malawi. *Social Indicators Research*, 135(2), 457-478. <https://doi.org/10.1007/s11205-016-1502-3>
  23. Keddell, E. (2019). Algorithmic justice in child protection. Statistical fairness, social justice and the implications for practice. *Social Sciences*, 8(10). <https://doi.org/10.3390/socsci8100281>
  24. Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles* [Computer software manual]. <https://www.tidymodels.org>
  25. Laderchi, C. R., Saith, R., & Stewart, F. (2003). Does it matter that we do not agree on the definition of poverty? A comparison of four approaches. *Oxford Development Studies*, 31(3), 243-274. <https://doi.org/10.1080/1360081032000111698>.
  26. Lepenies, P. (2020). Metrics, politics and definitions. How poverty lost its social context and what this means for current debates. En V. Beck, H. Hahn, & R. Lepenies (eds.), *Dimensions of poverty: Measurement, epistemic injustices, activism* (pp. 57-73). Springer International Publishing. <https://doi.org/10.1007/978-3-030-31711-94>.
  27. Lepri, B., Staiano, J., Sangokoya, D., Letouze, E., & Oliver, N. (2017). The tyranny of data? the bright and dark sides of data-driven decision-making for social good. *Transparent data mining for big and small data* (pp. 3-24). Springer International Publishing. [https://doi.org/10.1007/978-3-319-54024-5\\_1](https://doi.org/10.1007/978-3-319-54024-5_1)
  28. Lucci, P., Bhatkal, T., & Khan, A. (2018). Are we underestimating urban poverty? *World Development*, 103, 297-310. <https://doi.org/10.1016/j.worlddev.2017.10.022>
  29. Maccio, J., & Mitchell, A. (2023). Medición multidimensional de pobreza en ciudades segregadas. Evidencia de la ciudad de Buenos Aires. *Revista Desarrollo y Sociedad*, 93, 101-137. <https://doi.org/10.13043/DYS.93.3>
  30. Mahamallik, M., & Sahu, G. B. (2011). Identification of the poor. Errors of exclusion and inclusion. *Economic and Political Weekly*, 46(9), 71-77.
  31. Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28, 92-122.
  32. Putnam, D. (2020). Poverty as a social relation. En V. Beck, H. Hahn, & R. Lepenies (eds.), *Dimensions of poverty. Measurement, epistemic injustices, activism* (pp. 41-55). Springer International Publishing. <https://doi.org/10.1007/978-3-030-31711-93>.
  33. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., & Barnes, P. (2020). Closing the AI accountability gap. Defining an end-to-end framework for internal algorithmic auditing. En *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44). <https://doi.org/10.48550/arXiv.2001.00973>
  34. Roelen, K. (2017). Monetary and multidimensional child poverty: A contradiction in terms? *Development and Change*, 48(3), 502-533.
  35. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3-29.
  36. Sen, A. (1985). *Commodities and capabilities*. Oxford University Press.
  37. Sen, A. (1992). *Inequality reexamined*. Harvard University Press.
  38. Sen, A. (1999). *Development as freedom*. Anchor Books.
  39. Sen, A. (2018). *Collective choice and social welfare*. Harvard University Press.
  40. Sohnesen, T. P., & Stender, N. (2017). Is random forest a superior methodology for predicting poverty? An empirical assessment. *Poverty & Public Policy*, 9(1), 118-133.
  41. Tran, V. Q., Alkire, S., & Klasen, S. (2015). Static and dynamic disparities between monetary and multidimensional poverty measurement. Evidence from Vietnam. En T. I. Garner & K. S. Short (eds.), *Measurement of poverty, deprivation, and economic mobility* (vol. 23, p. 249-281). Emerald Publishing Ltd.
  42. United Nations Development Programme (UNDP) & Oxford Poverty and Human Development Initiative –OPHI (2021). *Global multidimensional poverty index 2021 – Unmasking disparities by ethnicity, caste and gender*. United Nations Development Programme and Oxford Poverty and Human Development Initiative. <https://ophi.org.uk/wp-content/uploads/UNDPPOPHIGMPI2021ReportUnmasking.pdf>
  43. United Nations Development Programme (UNDP) & Oxford Poverty and Human Development Initiative –OPHI. (2022). *2022 global multidimensional poverty index (MPI)*. Autor.
  44. United Nations –UN. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. *Transforming our world: The 2030 Agenda for*

- Sustainable Development*. Autor. <https://unctad.org/system/files/official-document/ares70d1en.pdf>
45. Usmanova, A., Aziz, A., Rakhmonov, D., & Osamy, W. (2022a). Utilities of artificial intelligence in poverty prediction. A review. *Sustainability*, 14(21). <https://doi.org/10.3390/su142114238>
  46. Usmanova, A., Aziz, A., Rakhmonov, D., & Osamy, W. (2022b). Utilities of artificial intelligence in poverty prediction. A review. *Sustainability*, 14(21), 14238.
  47. Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Fam med*, 37(5), 360-363.
  48. Villatoro, P., & Santos, M. E. (2019). ¿quiénes son pobres? Análisis de su identificación en América Latina. *Problemas del Desarrollo*, 50(199), 3-29.
  49. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Francois, R., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

## Anexos

### Anexo 1. Cálculo del IPM con la base de datos del Sisben IV

A continuación, se presentan los detalles para cada indicador con algunas reflexiones sobre el cálculo oficial cuando se alteran las mediciones (Experimento 3). En los casos donde no se alteró el cálculo, aparece la definición del indicador de acuerdo con la metodología oficial (Angulo *et al.*, 2016).

*Bajo logro educativo*. Se considera que un hogar está privado, cuando el promedio del logro educativo de las personas de quince años y más que lo componen es menor a nueve años escolares.

*Analfabetismo*. Se considera privados a los hogares donde al menos una persona de quince años o más no sabe leer y escribir. Para este cálculo, se bajó el puntaje de corte de la edad, se consideró analfabeta si una persona de ocho o más años no sabe leer ni escribir; en coherencia con las edades del sistema educativo para la enseñanza de lectura y escritura, aunque se dio un margen de dos años más, puesto que se espera que a los seis años la persona logre leer y escribir.

*Inasistencia escolar*. Se considera que un hogar está privado si menos del 100% de los niños de 6-16 años asiste al colegio.

*Rezago escolar*. Se considera que un hogar tiene privación en la variable si alguno de los niños de 7-17 años tiene rezago escolar.

*Servicios de primera infancia*. Este indicador se aproxima según las indicaciones del Sisben IV (cuadro 2). Entonces, se considera un hogar privado cuando tienen niños menores de cinco años sin seguridad social o en las siguientes categorías: “En casa solo”, “Al cuidado de un pariente menor de 18 años”, “Con su padre o madre en el trabajo”, de la pregunta del Sisben IV: “¿Dónde o con quién permanece... durante la mayor parte del tiempo entre semana (menores de 5 años)?”.

*Trabajo infantil*. Están en condición de trabajo infantil niños (5-17 años) que en su actividad principal en el último mes era: trabajando, buscando trabajo u oficios del hogar.

*Desempleo de larga duración*. Población en edad de trabajar (mayores de quince años) que se encuentran buscando trabajo.

*Trabajo informal*. En privación, cuando al menos una persona del hogar en edad de trabajar (mayor a quince años) está trabajando, pero no cotiza a pensión, o cuando el hogar no tiene población económicamente activa (trabajando o buscando trabajo).

*Sin aseguramiento en salud*. Un hogar se encuentra en privación, si alguno de sus miembros no está asegurado en salud. Dado que en el indicador de acceso a servicios para el cuidado de la primera infancia se observa si los niños entre 0 y 5 años se encuentran afiliados al Sistema General de Seguridad Social en Salud (SGSSS). Este indicador se mide solo para la población mayor a cinco años.

*Barreras de acceso a servicios de salud dada una necesidad*. El hogar se encuentra privado si una o varias personas del hogar sufrieron una enfermedad el último mes y no acudieron a los servicios de salud, o los que acudieron no fueron atendidos. Los hogares donde las personas no sufrieron alguna enfermedad el último mes se consideran no privados. Es importante señalar que en este indicador no es clara la privación sobre si la persona no acude a los servicios de salud por motivos de restricciones de acceso, o bien por voluntad propia.

*Acceso a fuente de agua mejorada*. Privación para hogares en área urbana si el hogar no cuenta con conexión

a servicio público de acueducto en la vivienda. Para la zona rural, se considera privado el hogar si obtienen el agua para preparar los alimentos de un pozo sin bomba, agua lluvia, río, quebrada, manantial, pila pública, carrotanque, aguatero o agua embotellada.

*Eliminación de excretas.* De acuerdo con la metodología, se consideran privados los hogares del área urbana sin conexión a servicio público de alcantarillado. En el área rural se encuentran en privación los hogares que tienen inodoro sin conexión, letrina o bajamar, o simplemente no cuentan con servicio sanitario.

*Pisos.* Se consideran privados los hogares con pisos de tierra.

*Paredes exteriores.* El hogar se encuentra privado en la zona urbana, cuando el material de las paredes exteriores del hogar es madera burda, tabla, tablón, guadua,

otro vegetal, chapa de zinc, tela, cartón, desechos o no tiene paredes. En la zona rural, el hogar está privado siempre que el material de las paredes exteriores sea guadua, otro vegetal, chapa de zinc, tela, cartón o desechos; o bien cuando no tiene paredes.

*Hacinamiento crítico.* Se considera que existe hacinamiento en el hogar (por tanto, privación), cuando el número de personas por cuarto para dormir, excluyendo cocina, baño y garaje, es mayor o igual a tres en el área urbana y de más de tres personas por cuarto para la zona rural.

## Anexo 2. Dimensiones, privaciones y pesos del IPM

La tabla siguiente contiene el nombre de la dimensión, la privación y los pesos de la metodología oficial y los del Experimento 2.

| Dimensión  | Privación   | Pesos del IPM Colombia | Pesos del Experimento 2 |
|--|---|------------------------|-------------------------|
| Condiciones educativas del hogar                                       | Bajo logro educativo  | 0.10                   | 1/15                    |
|  | Analfabetismo   | 0.10                   | 1/15                    |
| Condiciones de la niñez y juventud                                     | Inasistencia escolar  | 0.05                   | 1/15                    |
|  | Rezago escolar  | 0.05                   | 1/15                    |
|  | Barreras de acceso a servicios para el cuidado de la primera infancia | 0.05                   | 1/15                    |
|  | Trabajo infantil  | 0.05                   | 1/15                    |
| Trabajo  | Desempleo de larga duración   | 0.10                   | 1/15                    |
|  | Empleo informal   | 0.10                   | 1/15                    |
| Salud  | Sin aseguramiento en salud  | 0.10                   | 1/15                    |
|  | Barreras de acceso a servicio de salud                                | 0.10                   | 1/15                    |
| Acceso a servicios públicos domiciliarios y condiciones de la vivienda | Sin acceso a fuente de agua mejorada                                  | 0.04                   | 1/15                    |
|  | Inadecuada eliminación de excretas                                    | 0.04                   | 1/15                    |
|  | Pisos inadecuados   | 0.04                   | 1/15                    |
|  | Paredes exteriores inadecuadas  | 0.04                   | 1/15                    |
|  | Hacinamiento crítico  | 0.04                   | 1/15                    |

Fuente: elaboración propia.