

Análisis de Perfiles con Técnicas No Supervisadas de Minería de Datos como Soporte a la Toma de Decisiones en la Estrategia UNIDOS para la Superación de la Pobreza en Colombia

Andrés Eugenio SILVA MONSALVE

Escuela Colombiana de Ingeniería Julio Garavito
Bogotá, 111166/Cundinamarca, Colombia

Dante CONTI

Universidad Estatal de Campinas
Campinas, 13083-970/Estado de São Paulo, Brasil

Victoria Eugenia OSPINA BECERRA

Escuela Colombiana de Ingeniería Julio Garavito
Bogotá, 111166/Cundinamarca, Colombia

RESUMEN

En la búsqueda de garantizar que los hogares más pobres y vulnerables puedan superar sus condiciones de vida, la Estrategia Red UNIDOS, consolida y gestiona información que permite conocer la situación de los hogares y personas, para la toma de decisiones relacionadas con la planeación y ejecución del proceso de acompañamiento. Este proyecto determina el valor agregado y potencial de la información mediante modelado basado en datos, específicamente clustering. Al respecto, se crea una vista minable donde se aplica clusterización jerárquica (método Ward, con distancia coseno) y se decide cortar los dendogramas en tres (3) y seis (6) grupos respectivamente, que posteriormente fueron postprocesados y validados para describir en detalle los perfiles y su interpretación usando los centroides de cada grupo. Finalmente, los perfiles hallados y plenamente descritos sirven como base para dar soporte a la toma de decisiones al proceso de acompañamiento de la Estrategia Red UNIDOS y son resumidos a su vez en dos (2) dashboards de resultados.

Palabras Claves: Minería de datos, Clusterización, técnicas no supervisadas, pobreza, Estrategia Unidos.

1. INTRODUCCIÓN

Colombia calcula la pobreza a partir de los ingresos monetarios de los hogares y a través del Índice de Pobreza Multidimensional (IPM), el cual, evalúa privaciones en la educación de la niñez y juventud, el trabajo, la salud, el acceso a servicios públicos y las condiciones de vivienda. La pobreza en 2015 se calcula en 10,3 por ciento de la población y el gobierno a través de diferentes programas brinda ofertas para suplir las necesidades básicas de los hogares en condición vulnerable y de esta forma disminuir las cifras (Administrativo, Estadística, & Nacional, 2014) [1].

A través de la Ley 1785 de 2016, se establece la Red para la Superación de la Pobreza Extrema Red UNIDOS, dicha Red estará conformada por las entidades del Estado que presten servicios sociales dirigidos a la población en pobreza extrema, alcaldías y gobernaciones, el sector privado y organizaciones de la sociedad civil, que busca asegurar que los hogares más pobres y vulnerables puedan superar las condiciones que los mantienen

en pobreza y consoliden sus capacidades para el desarrollo y el ejercicio de sus derechos (Congreso de la República de Colombia, 2016) [2].

Para la implementación de la Estrategia Red UNIDOS, se consolida y se gestiona la información a través del Sistema de Información Misional, el cual, provee datos específicos, que permite conocer la situación de las comunidades, hogares y personas focalizadas, para la toma de decisiones relacionadas con la planeación, implementación y ejecución de la citada Estrategia.

En el acompañamiento familiar se levanta la información de caracterización del hogar y se evalúa las condiciones de vida de cada integrante que conforma el núcleo. A partir de los déficits detallados en el hogar se genera la demanda en una serie de dimensiones como son Identificación, Educación y Capacitación, Ingresos y Trabajo, Bancarización y Ahorro, Salud, Nutrición, entre otras, que deben ser cubiertas con ofertas de entidades públicas del estado y empresa privada.

Sin embargo, la gestión de información y la madurez para tomar decisiones dependen no solo del uso de tecnologías de punta, sino de la tendencia, filosofía o cultura que la organización debe cultivar al interior y exterior para generar apropiación y valor cuantitativo y cualitativo sobre cada uno de los datos que hay en el ambiente empresarial y social, además de la pertinencia en los datos internos y externos que responden preguntas y generan valor para la articulación sector social.

Para Davenport, Aclis y Prusak (1999) [3], la información se ha convertido en la "moneda" clave de la organización, se ha vuelto demasiado valiosa para la mayoría de los gerentes como para regalar. Para que las organizaciones basadas en la información tengan éxito, las empresas deben aprovechar el poder de la política, es decir, permitir que las personas negocien el uso y la definición de la información, del mismo modo que negociamos el intercambio de otras monedas.

Es así como hoy se observan políticas, directrices, planes de gobierno o estratégicos, planes de acción, indicadores, estrategias de ventas y mapas de riesgos, con mejoras importantes no solo con antecedentes y experiencia del negocio sino con otras variables que permiten ser más asertivos en su funcionamiento de los procesos del negocio. Lo anterior con el fin de generar valor y sostenibilidad en el sector social y la población atendida.

2. FUNDAMENTACIÓN TEÓRICA

Iniciativas de medición de Pobreza:

Para Alkire & Foster (2007) [4], actualmente existen tres enfoques principales para identificar a los pobres en un contexto multidimensional. El primer enfoque es el 'unidimensional', a través del cual se combinan los distintos indicadores de bienestar en una sola variable agregada y una persona es identificada como pobre cuando la variable cae debajo de una determinada línea de corte.

Para Alkire et al (2007) [4], el segundo es el enfoque de 'unión', que considera a una persona que sufre privaciones en una sola dimensión como pobre en el sentido multidimensional. Generalmente se reconoce que este enfoque es excesivamente inclusivo y puede llegar a generar estimaciones exageradas de la pobreza. El tercer enfoque principal es el método de la 'intersección', que exige que una persona sufra privaciones en todas las dimensiones para ser identificada como pobre.

Para Angulo, Díaz & Pardo (2011) [5], el Departamento Nacional de Planeación permite utilizar la metodología Alkire y Foster (2007) como instrumento de seguimiento y monitoreo de las acciones de política pública para la reducción de la pobreza. La nitidez de la noción multidimensional de la pobreza que expresa el indicador se transmite a la discusión multisectorial del diseño y la planeación de estrategias para la reducción de la pobreza. Es posible generar alarmas a partir de la identificación de las dimensiones y variables con mayor privación, las más rezagadas y las más críticas en la población pobre.

Para Prosperidad Social, (2015) [6] Colombia cuenta con dos (2) metodologías oficiales de medición de pobreza: Pobreza Monetaria y el Índice de Pobreza Multidimensional –IPM Colombia– (CONPES Social 150, 2012).

Cada una de estas metodologías aborda aspectos distintos pero de vida.

Minería de datos: Como concepto para Microsoft (2018), la minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

Técnicas no supervisadas: De acuerdo con Moreno & López (2018) [8], las técnicas de minería de datos no supervisadas, también conocidas con el nombre de técnicas de descubrimiento del conocimiento, se utilizan para la detección de patrones ocultos en bases de datos de gran tamaño. Dichos patrones representan por sí mismos información útil que puede ser utilizada directamente en la toma de decisiones. Trabajos recientes muestran que los algoritmos no supervisados pueden utilizarse con éxito para resolver problemas de clasificación.

Al respecto, para Foster & Fawcett (2013) [9], los métodos jerárquicos se dividen en:

“Los métodos jerárquicos aglomerativos comienzan con tantos clústers como objetos tengamos que clasificar y en cada paso se recalculan las distancias entre los grupos existentes y se unen los dos grupos más similares o menos disímiles. El algoritmo acaba con un clúster conteniendo todos los elementos.

Los métodos jerárquicos divisivos comienzan con un clúster que engloba a todos los elementos y en cada paso se divide el grupo más heterogéneo. El algoritmo acaba con tantos

complementarios del fenómeno de la pobreza y son los indicadores principales para el seguimiento a las políticas sociales incluidas en el último Plan Nacional de Desarrollo.

Para Prosperidad Social, (2015) [7] en la guía del acompañamiento familiar, en el país también se ha reconocido la pobreza como un fenómeno multidimensional, que aborda no solo aspectos monetarios, sino que tiene en cuenta la situación de los hogares en 15 indicadores de bienestar relacionados con: i) educación; ii) niñez y la juventud; iii) trabajo; iv) salud; y v) servicios públicos domiciliarios. La pobreza multidimensional es entendida como una situación en la cual un hogar tiene alrededor de cinco (5) carencias en quince (15) indicadores seleccionados. En el PND 2014-2018 se estableció como meta la reducción de la pobreza multidimensional de 24,8% en 2013 a 17,8% de 2018. Estos dos enfoques en la conceptualización de la pobreza implican reconocer tanto la perspectiva monetaria como multidimensional en el proceso de superación de la pobreza.

Prosperidad Social, (2015) [6], para garantizar el Acompañamiento Familiar y Comunitario, y el acceso a los servicios Sociales del Estado dirigido a los hogares en pobreza extrema previstos por la Ley 1785 de 2016, Prosperidad Social implementa, como parte de su oferta social, la Estrategia UNIDOS, brinda Un Acompañamiento Familiar a los hogares en pobreza extrema a fin de lograr que cada uno de ellos reconozca sus fortalezas y potencialidades, consolide sus vínculos familiares, sus redes de interacción social, adquiera o afiance habilidades sociales y acceda a la oferta de bienes y servicios institucionales, para superar su situación de pobreza. Este acompañamiento tiene metodologías y tiempos definidos orientados a que los beneficiarios alcancen condiciones mínimas

clústers (de un elemento cada uno) como objetos se hayan clasificado.”

Algunos de los métodos jerárquicos aglomerativos para Moreno (2018) [8] son método del enlace simple, método de enlace completo, método del promedio entre grupos, método del centroide, método de la mediana y método de Ward.

Clusterización y distancia por similitud: Para Han, Kamber & Pei. (2012) [10], las técnicas de agrupamiento consideran las tuplas de datos como objetos. Particionan los objetos en grupos, o clústers, para que los objetos dentro de un clúster sean "similares" entre sí y "disímiles" a objetos en otros grupos. La similitud se define comúnmente en términos de cómo "Cerrar" los objetos están en el espacio, en función de una función de distancia. La "calidad" de un clúster puede representarse por su diámetro, la distancia máxima entre dos objetos cualquiera en el grupo.

Para Foster et al (2013) [9], los métodos jerárquicos tienen por objetivo agrupar clústeres para formar un nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o bien se maximice alguna medida de similitud.

Para Lima & Boccado (2014) [11], el método de coeficiente de correlación de Spearman, muestra una asociación entre variables. Permite obtener un coeficiente de asociación ente variables que no se comportan normalmente, entre variables ordinales. Se calcula con base a una serie de rangos asignados. los valores van de 1 a 1, siendo 0 el valor que indica no correlación, y los signos indican correlación directa e inversa.

Para Han, et al. (2012) [10], el método de Ward es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clústeres para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, dentro de cada clúster, de cada individuo al centroide del clúster.

Al respecto Foster et al. (2013) [9], el objetivo del método de Ward es encontrar en cada etapa aquellos dos clústeres cuya unión proporcione el menor incremento en la suma total de errores.

Todos los métodos de clusterización no supervisados tienen en común que para llevar a cabo agrupaciones necesita definir y cuantificar la similitud. Al respecto, para De la Fuente (2011) [12] poder unir variables o individuos es necesario tener algunas medidas numéricas que caractericen las relaciones entre las variables o los individuos. Cada medida refleja asociación en un sentido particular y es necesario elegir una medida apropiada para el problema concreto que se esté tratando. La medida de asociación puede ser una distancia o una similitud.

Respecto a medidas de distancia, para Amat (2017) [13], el término distancia se emplea dentro del contexto del clustering como cuantificación de la similitud o diferencia entre observaciones. Si se representan las observaciones en un espacio p dimensional, siendo p el número de variables asociadas a cada observación, cuando más se asemejen dos observaciones más próximas estarán, de ahí que se emplee el término distancia.

3. ¿QUÉ PROBLEMA SE EVIDENCIA?

Bajo las premisas anteriores, esta propuesta de investigación se focaliza en determinar el valor agregado y potencial de la información mediante modelado y análisis de datos, de tal manera que se facilite el procesamiento e interpretación de ésta y así, dar soporte optimizado al proceso de toma de decisiones asociado a la evaluación y clasificación de perfiles en la estrategia UNIDOS para la superación de la pobreza en Colombia.

4. ¿CÓMO SE ANALIZÓ?

La investigación se basa en la metodología Knowledge Discovery in Databases (KDD) introducida por Fayyad [14] en 1996 y a través del preprocesamiento de los datos mediante informes exploratorios y de estadística descriptiva que tiene como fuente de información las bases de datos de los hogares acompañados en Colombia que se encuentran en el sistema de información misional de la Estrategia Unidos de Prosperidad Social, caracterizadas por contar con un gestor social, el cual realiza el acompañamiento. Estos hogares cuentan con información de ubicación actualizada en los años 2016 y 2017, además de contar con el cálculo de los veintiséis (26) indicadores de pobreza denominados logros con detalle familiar e individual. La diversidad en información respecto al avance del acompañamiento y las situaciones en la condición de vida de cada hogar será útil para relacionar las variables objeto indispensable en la toma de decisiones.

Para la selección de la muestra se tomará como referente los hogares con acompañamiento familiar de la Estrategia Unidos en los departamentos de Caldas, Quindío y Risaralda que se encuentren registrados en el sistema de información misional. Con base en los criterios anteriores el tamaño aproximado para el procesamiento de datos corresponde a 25.464 registros.

5. ANÁLISIS DESCRIPTIVO

Se presenta una visión global del conjunto de datos de los hogares pobres que son objeto del análisis, en el cual, se detalla la distribución de los integrantes que conforman los hogares por quinquenio y sexo. Se destaca la distribución porcentual por estado de cada uno de los logros que calcula la Estrategia Red UNIDOS previa caracterización de los hogares pobres. Es importante acotar que la línea de investigación es la búsqueda de agrupamientos de acuerdo con los comportamientos de la información. Para el caso descriptivo, este agrupamiento se realiza desde la perspectiva del negocio y se concluye con una propuesta de medición de hogares de acuerdo con la dimensión de logros y la interpretación de los grupos producto del ejercicio.

Para iniciar el ejercicio se toma como fuente de información la base de datos que comprende una población de noventa y ocho mil novecientos cincuenta y un (98.951) personas agrupadas en veinticinco mil cuatrocientos sesenta y cuatro (25.464) hogares, acompañados durante los años 2016 y 2017 en los departamentos de Caldas, Quindío y Risaralda (eje cafetero). A continuación, se detalla el análisis a través de tres agrupamientos objetivo, así:

Pirámide poblacional analizada: Población muy homogénea por sexo, y con altas concentraciones en los grupos de adolescencia y adulto joven al desagregarse en grupos quinquenales, ver figura 1. El comportamiento en los grupos quinquenales en la población pobre analizada tiene una alta similitud con la pirámide poblacional del censo 2015 realizado por el Departamento Administrativo Nacional de Estadística (DANE), para cada uno de los departamentos, en donde se visualiza también altas concentraciones de personas en la población adolescente y joven.

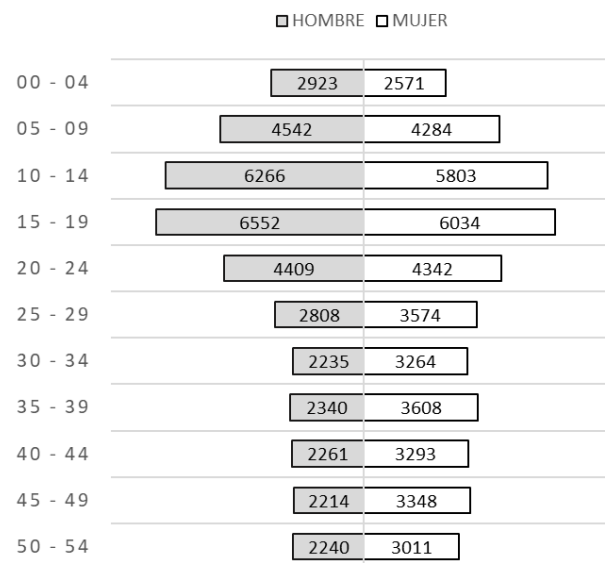


Figura 1 Cantidad de personas por género y quinquenio

La Estrategia Unidos establece para un hogar identificado como pobre, veintiséis (26) indicadores o escenarios que deben predominar para mejorar su condición, distribuidos en cinco (5) dimensiones.

A través de algoritmos por cada indicador y con la información previamente recolectada en el formulario de caracterización, se calcula y asigna un valor a cada indicador (Alcanzado, Por Alcanzar o No Aplica), el cual, permite una acción o gestión particular por parte de la Estrategia o sus aliados y permite cuantificar el avance del hogar en términos de superación de pobreza.

Indicadores familiares de los hogares acompañados: La Estrategia Unidos establece para un hogar identificado como pobre, veintiséis (26) indicadores o escenarios que deben predominar para mejorar su condición, distribuidos en cinco (5) dimensiones.

A través de algoritmos por cada indicador y con la información previamente recolectada en el formulario de caracterización, se calcula y asigna un valor a cada indicador (Alcanzado, Por Alcanzar o No Aplica), el cual, permite una acción o gestión particular por parte de la Estrategia o sus aliados y permite cuantificar el avance del hogar en términos de superación de pobreza.

Para el análisis de correlación (método Spearman) de los indicadores se homogeniza las variables tomando únicamente la frecuencia del estado Por Alcanzar (déficit del hogar) por municipio, ver figura 2. Al respecto, se evidencian varias relaciones positivas y se agrupan los indicadores con mayor peso en tres (3) grupos. A (Indicador 7 y 8), B (Indicador 9, 10 y 19), C (Indicador 16, 19 y 20). En consecuencia, los grupos A y B tienen indicadores que son requeridos para el proceso de superación de pobreza. En el caso A- “Educación infantil”, un menor de edad sin acceso al sistema educativo es más propenso a ejercer trabajo infantil. En el caso B- “Acceso a servicios”, un hogar con acceso a servicios básicos como agua y saneamiento básico tiene mayor probabilidad de acceder a herramientas digitales. Y en el caso C- “Responsabilidad”, en donde una persona con educación en derechos sexuales y reproductivos también tienen educación financiera y utilización de herramientas digitales. Es importante indicar que también se visualizan relaciones con peso negativo que para el caso no son clasificadas dado que no representan valor en el negocio. Estas relaciones negativas son (Indicadores 18 y 26), (Indicadores 12 y 26), (Indicadores 19 y 26), (Indicadores 2 y 18).

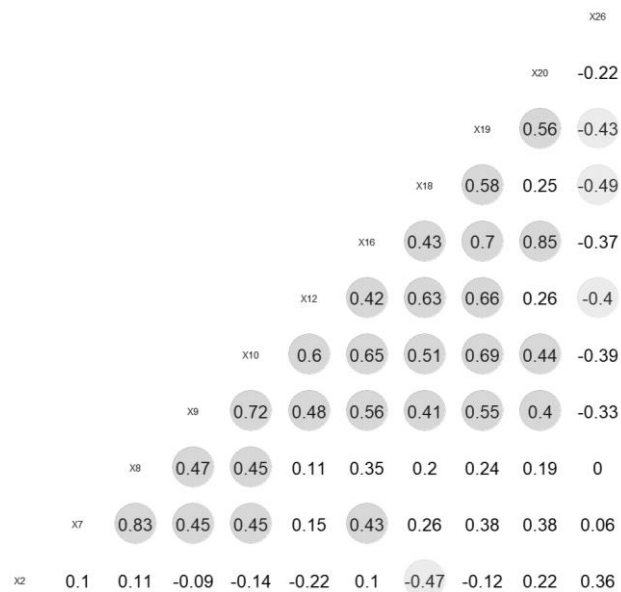


Figura 2 Pirámide Correlación logros familiares Spearman

Un segundo ejercicio con los indicadores familiares consiste en agruparlos de acuerdo con la similitud en los estados adquiridos para la cobertura municipal analizada. De este modo se identifican los grupos **D-Debilidades**, **E-Baja Aplicabilidad** y **F-Fortalezas**.

Grupo Debilidades. Indicadores (18, 19 y 25), relacionado con estudios postsecundarios, herramientas digitales y vinculación laboral. En este grupo se asocian aquellos indicadores de difícil cumplimiento a corto plazo para un hogar pobre. Se visualiza con claridad una estrecha relación entre una vinculación laboral formal y el desarrollo de habilidades para el trabajo.

Grupo de baja aplicabilidad. Indicadores (3, 4, 5, 6, 13 y 14), relacionado con primera infancia y discapacidad. En este grupo se asocian aquellos indicadores que por lo regular No Aplican para el hogar dado que no cuenta con integrantes en situación de discapacidad o en grupo etario de primera infancia.

Grupo Fortalezas. Indicadores (1, 2 y 21) relacionado con identificación, aseguramiento de salud y pisos de la vivienda. En este grupo se asocian aquellos indicadores de fácil cumplimiento.

Indicadores individuales de las personas acompañadas: Las características de dieciocho (18) indicadores, la Estrategia calcula inicialmente la calificación a nivel de la persona. A partir de esta información se ponderan para conocer la calificación familiar entre todos los integrantes del hogar. A partir del detalle de la calificación individual se realiza el ejercicio de correlación (método Spearman) y se confirma la relación del grupo A- “Educación infantil” y D- “Debilidades”, previamente identificados con los ejercicios anteriores. Sin embargo, al verificar las relaciones negativas con mayor peso, se identifica un nuevo grupo G (Indicador 8 y 17). En este caso G- “Educación y Trabajo” hay una relación inversamente proporcional entre saber leer y escribir con el trabajo infantil.

Si bien los indicadores se encuentran agrupados por dimensiones que distan de su objeto, si se evidencia interacciones al compararlos, encontrando siete (7) grupos focales en donde los indicadores apoyan la lectura o comportamiento del conjunto. Los grupos focales son:

- A. Educación Infantil
- B. Acceso a servicios
- C. Responsabilidad
- D. Debilidades
- E. Baja aplicabilidad
- F. Fortalezas
- G. Educación y Trabajo

El comportamiento de la muestra de indicadores familiares e individuales es reflejo de los porcentajes a nivel nacional y municipal. Se resalta que la comparación de indicadores familiares e individuales es igual, dado que uno se construye a partir de la ponderación del otro. Al respecto el enfoque del proyecto será únicamente a partir de los indicadores familiares.

Se evidencia que por dimensión no hay un dato que permita el análisis por este agrupamiento y que refleje o resuma el estado interno de sus indicadores. Producto del análisis se validan los pesos por dimensión a partir del Índice de Pobreza Multidimensional (IMP) y se distribuye el sub-peso de los indicadores por cada dimensión asignando el 80% a los requeridos y el 20% para los deseados, cuando el estado del indicador es Alcanzado o No aplica.

6. ANÁLISIS EXPLORATORIO

Presenta los resultados del postprocesamiento de hogares pobres que son objeto del análisis a través de las herramientas de clusterización jerárquica (método Ward, con distancia coseno) y se decide cortar los dendogramas en tres (3) y seis (6) grupos respectivamente, que posteriormente fueron interpretados y validados para describir en detalle los perfiles usando los

centroides de cada grupo y las variables principales.

Aplicación de técnicas no supervisadas: Se realiza la clusterización usando el método WARD a razón que garantiza la mínima varianza con clústeres más compactos que permiten ser más fáciles de interpretar. Y a través de los algoritmos de distancia por similitud de coseno, euclídea y manhattan, se forman las etiquetas o clasificadores realizando cortes para formar tres (3) y seis (6) grupos respectivamente. Estos cortes (k=3 y k=6) se grafican a través de dendogramas para visualizar el ejercicio.

Finalizada la diagramación por dendogramas y para continuar con el análisis exploratorio, se limita a la similitud coseno porque es muy útil en datos con cambios de signo como se puede observar en los valores que toman las variables de los logros que para el caso es 1, 0, -1. Es importante resaltar que la similitud por coseno mide esa direccionalidad en los datos, además de ofrecer un ángulo que puede diferenciar los valores tomados. Otras distancias por similitud como manhattan o euclidean anulan ese sentido porque al elevarse al cuadrado dan positivas.

Respecto al etiquetado se realiza en cada uno de los hogares objeto del análisis y por cada etiqueta creada en la fuente de información a partir del análisis exploratorio de las similitudes mencionados con los cortes en k=3 y k=6, se efectúa el cálculo de la media, mediana, desviación estándar, máximo y mínimo.

Este resultado consolidado permite que a través de Random Forest verificar la integridad, acertando en un 96.5% (tan solo 3.5% de error) en la clasificación de los hogares para el etiquetado por tan solo tres grupos, sin embargo, al aumentar a una clasificación por seis grupos se disminuye a un 92% de asertividad (con un error del 8%).

Priorización de variables: El Random Forest permite identificar la importancia de las variables que definen la variedad y tipología de la muestra cuando el clasificador es para tres grupos.

Para el etiquetado por tres grupos, el diagrama de Pareto de la figura 3, muestra las variables a tener en cuenta para el análisis de acuerdo con aquellas que presentan mayor peso respecto al porcentaje acumulado, para el análisis llama la atención el Score5, Score4 (hace referencia a las dimensiones Ingresos y Trabajo y Habitabilidad respectivamente) y la variable LOGRO_03 que es parte integral del cálculo del Score2 (dimensión Salud y Nutrición), sin embargo, para tener mayor detalle se tomarán las variables siguientes LOGRO_12 y LOGRO_06 que hace referencia respectivamente a la construcción del Score1 y Score3 (dimensiones de Identificación y Educación y Capacitación) y de esta forma se complementa el universo de las dimensiones mencionadas.

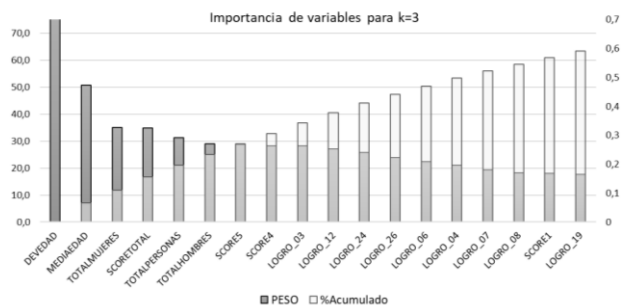


Figura 3 Importancia de variables para clasificación por tres grupos

Interpretación de perfiles: A continuación, se sintetiza en tabla el análisis del agrupamiento para k=3, el cual detalla los aspectos principales y características que brindan importancia a cada perfil y demuestra las diferencias al ser comparados entre ellos.

K=3	Etiqueta 1	Etiqueta 2	Etiqueta 3
Nombre propuesto para el grupo	PROYECTO FORMACIÓN	PROYECTOS PRODUCTIVOS	PROYECTO SALUD Y NUTRICIÓN
Porcentaje de marcación	30.85%	51.96%	17.19%
Promedio de edad en el hogar	23 años (+/- 5)	33 años	63 años
Promedio de integrantes en el hogar	5 (+/- 2)	4 (+/- 1)	1
Características del grupo	Hogar conformado mayoritariamente por mujeres jóvenes. Cuentan con condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado laboral.	Hogar conformado equitativamente por género. Con tendencia a la ausencia de primera infancia y niños. Tienen condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado.	Hogar con tendencia unifamiliar de adultos mayores. Cuentan con condiciones apropiadas en los escenarios habitacionales y de ingreso. Pero con una notable ausencia en condiciones de salud y nutrición.
Posible sugerencia	Acercar ofertas de estudio como media técnica, tecnológica y/o universitaria.	Acercar ofertas de estudio como tecnológica y/o universitaria. Adicional de cursos complementarios para incentivar el emprendimiento e incubación de empresa.	Oferta de promoción y prevención en salud y nutrición

Para conceptualizar el ejercicio con los resultados del postprocesamiento y tomar decisiones a través de la lectura desde un visualizador que brinde mayor eficacia y eficiencia en la interpretación de los perfiles.

7. CONCLUSIONES

En esta investigación se ha explotado la información recolectada por la Estrategia Unidos de los hogares pobres focalizados por Prosperidad Social en los departamentos de Caldas, Quindío y Risaralda. El conjunto de datos analizados adopta la metodología del Índice de Pobreza Multidimensional a través del estado de los veintiséis logros familiares que se distribuyen en cinco dimensiones

El análisis descriptivo a través de la correlación de variables

permitió conocer la composición demográfica y la situación de los hogares, agrupados en siete subconjuntos con características comunes en el estado de logros (Educación Infantil, Acceso a servicios, Responsabilidad, Debilidades, Baja aplicabilidad, Fortalezas, Educación y Trabajo). Los subconjuntos detallados corresponden al primer objeto de intervención social por parte del estado colombiano a través de entregas de beneficios tangibles e intangibles para el mejoramiento y permanencia de las condiciones.

Para comparar los resultados proporcionados por el análisis descriptivo, se asumió el cálculo del score del hogar a través de una distribución equitativa por dimensión, pero internamente en cada dimensión con sub-pesos diferenciados por la importancia del logro para la superación de la pobreza de acuerdo con la metodología de la Estrategia Unidos.

Las variables existentes junto con las nuevas variables score hicieron parte del análisis exploratorio a través de la identificación de etiquetas por medio del cálculo de distancias por similitud (técnicas no supervisadas). El ejercicio permite conocer tres diferentes perfiles de hogares que se pueden detallar hasta seis tipos, además de las variables que tienen mayor importancia para la creación del agrupamiento, de las cuales se destacan el tipo score que tienen un análisis interno más amplio dada su constitución.

Clasificar un hogar por alguno de los tres perfiles tiene un error de tan solo 3.5% y normalmente sucede en la intersección del promedio de edad del hogar en dos de los perfiles identificados.

Los perfiles identificados con previa interpretación del negocio y visualización, permite tomar decisiones de tipo estratégico, táctico y operativo respecto al acompañamiento e intervención posterior de un hogar.

Tanto para el análisis descriptivo como para el exploratorio/resultados del postprocesamiento del clustering, se construye una representación gráfica que comprende las variables del negocio (ubicación geográfica, estado de logros y ciclo vital de hogares) que se complementa con las variables producto de la investigación (cálculos score y agrupamientos sugeridos con cortes en tres y seis). Esta visualización permite al usuario explorar la información de forma amigable, comprender el comportamiento y dinámicas de los hogares a nivel departamental y municipal, conocer la distribución por sexo y ciclo vital, responder preguntas de la operación del acompañamiento familiar que brinda Prosperidad Social a través de la Estrategia Red UNIDOS en los procesos de caracterización, plan de trabajo familiar, gestión oferta orientado a la superación de pobreza y la consecución de los logros que los hogares tengan por alcanzar o construcción de temáticas para fortalecer los logros que los hogares tengan cumplidos. Finalmente tomar decisiones a través de la lectura del visualizador con mayor eficacia y eficiencia a través de los perfiles interpretados.

En consecuencia, esta nueva experiencia de intervención mejora la prestación del servicio a la población vulnerable, permitiendo estandarizar a través de la identificación, planeación, ordenamiento, priorización y garantía del uso adecuado de los recursos del estado colombiano.

8. REFERENCIAS

[1] Departamento Administrativo Nacional de Estadística en Colombia (2018). Pobreza Monetaria y Multidimensional en Colombia. Boletín Técnico.

https://www.dane.gov.co/files/investigaciones/...vida/pobreza/bol_pobreza_17.pdf

[2] Congreso de la República de Colombia. (2016). Ley 1785 de 21 de junio de 2016 “Por medio de la cual se establece la Red para la Superación de la Pobreza Extrema Red Unidos y se dictan otras disposiciones,” 1–8. Retrieved from http://es.presidencia.gov.co/normativa/normativa/LEY_1785_DEL_21_DE_JUNIO_DE_2016.pdf

[3] Davenport, T, Eccles, R, & Prusak, L (1999). Chapter 2 - Information Politics, In Knowledge and Special Libraries, 29-48. <https://doi.org/10.1016/B978-0-7506-7084-5.50004-X>

[4] Sabina, A & Foster, J. Recuento y medición multidimensional de la pobreza. Documento de trabajo OPHI No. 7

[5] Angulo, R., Diaz, Y., & Pardo, R. (2011). Índice de Pobreza Multidimensional para Colombia (IPM-Colombia) 1997-2010. Departamento Nacional de Planeación. <https://colaboracion.dnp.gov.co/cdt/estudios%20economicos/382.pdf>

[6] Departamento para la Prosperidad Social (2017). Guía del Acompañamiento Familiar.GGA2. Versión 2. Grupo de Diseño Metodológico y Formación.

[7] Departamento para la Prosperidad Social (2017). Manual operativo de la estrategia de acompañamiento familiar y comunitario. MGA1 Versión 5. Grupo de Diseño Metodológico y Formación.

[8] Moreno, M., & López, V. (2018). Uso de técnicas no supervisadas en la construcción de modelos de clasificación en ingeniería de software. www.lsi.us.es/redmidas/Capitulos/LMD14.pdf

[9] Foster, P., Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking

[10] Jiawei Han, Micheline Kamber & Jian Pei. (2012). 3 - Data Preprocessing, In the Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, Boston, Pages 83-124, Data Mining (Third Edition), ISBN 9780123814791, <https://doi.org/10.1016/B978-0-12-381479-1.00003-4>.

[11] Lizma, P. & Boccado, G. (2014). Guía de Asociación entre variables (Pearson y Spearman en SPSS). Universidad de Chile. https://www.u-cursos.cl/facso/2014/2/SO01007/1/material_docente/bajar?id_material=994690.

[12] De la Fuente, S. (2011). Análisis de conglomerados. Universidad Autónoma de Madrid. <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf>

[13] Amat, J. (2017). Clustering y heatmaps: Aprendizaje No Supervisado. https://rpubs.com/Joaquin_AR/310338

[14] Fayyad, U., Stolorz, P. (1997) Data mining and KDD: Promise and challenges, Future Generation Computer Systems, 99-115. [https://doi.org/10.1016/S0167-739X\(97\)00015-0](https://doi.org/10.1016/S0167-739X(97)00015-0)

[15] Earley, S., Henderson, D., & Data Management Association. (2017). DAMA-DMBOK: Data management body of knowledge.

[16] Consejo Nacional de Política Económica y Social de Colombia. (2012). Conpes Social 150. Departamento Nacional de Planeación. <https://colaboracion.dnp.gov.co/CDT/Conpes/Social/150.pdf>