Universidad de los Andes

Maestría Economía Aplicada: *Big Data and Machine Learning for Applied Economics*

**Grupo:** Laura Natalia Capacho, Sebastián David Beltrán y Yurani Gonzalez[[1]](#endnote-1)

**GitHub URL:** <https://github.com/sbeltro/G10_PS2>

***Problem Set 2: Predicting Poverty[[2]](#footnote-1)***

**1. Introducción:**

En Colombia, el índice de pobreza monetaria fue de 39,3% en 2021 y la pobreza multidimensional de 16,0% (DANE, 2022). La ONU (2022) señala la erradicación de la pobreza como un imperativo ético, social, político y económico a nivel mundial, por esto, identificar correctamente los hogares que deben priorizarse es fundamental. El objetivo de este trabajo es construir un modelo predictivo de pobreza en hogares colombianos. Se caracterizó la pobreza desde dos enfoques: un problema de clasificación, donde encontramos que 25,7% de hogares en la base de prueba son pobres, y un modelo de predicción de ingreso comparado con la línea de pobreza, que indica que 32,1% de dichos hogares son pobres.

Destacamos que durante el estudio encontramos que las variables disponibles en la fuente de datos dispuesta para el ejercicio[[3]](#endnote-2), no permiten una aproximación sobresaliente de pobreza, pues las características del hogar que contiene son limitadas, y no se encuentran variables correspondientes a características de vivienda o condiciones de vida, comúnmente usadas para identificar pobreza (Kambuya, 2020). Para lograr una mejor caracterización de los hogares y enfocar de forma óptima las intervenciones es necesario ampliar las preguntas que contiene la encuesta para capturar otros factores, como lo hace la ELCA[[4]](#endnote-3).

**2. Datos**

En el estudio se utilizaron 4 bases[[5]](#endnote-4) de la GEIH 2018. En la tabla 1 se presentan las principales estadísticas descriptivas de las variables de interés con las que se entrenaron los modelos. Tenemos una muestra de 164,960 hogares, 33,024 clasificados como pobres y 131,936 no pobres. Se observa que en todas sus características existe una diferencia significativa entre los grupos: una mayor proporción de hogares pobres recibe subsidios (82,2% frente al 46% de no pobres), y ayudas monetarias de hogares nacionales e instituciones (29,6% y 31,6% respectivamente, frente al 20,6% y 11,0% en no pobres). Además, vemos que los hogares pobres duermen en promedio 2 personas por cuarto, el doble que en los no pobres.

En la tabla 2 se caracterizan los hogares que buscamos clasificar (*test).* La muestra se compone de 66,168 hogares, de los cuales, el 56,2% recibe subsidios, 13,2% subsidio familiar y tan solo 0,2% educativo. Asimismo, encontramos que el 23,2% recibió ayuda de hogares nacionales y 15,7% de instituciones, y vemos que en esta muestra duermen cerca de 2 personas en cada cuarto. (En la Gráfica 1 se observan las características de las dos muestras).

1. **Notas**

   Los códigos de los estudiantes del grupo son 202121025, 202121021 y 201212100, respectivamente. [↑](#endnote-ref-1)
2. \* Todas las tablas y gráficas mencionadas en el documento se encuentran en la sección de Anexos al final. [↑](#footnote-ref-1)
3. (Gran Encuesta Integrada de Hogares 2018 del DANE). [↑](#endnote-ref-2)
4. Encuesta Longitudinal Colombiana de la Universidad de Los Andes. [↑](#endnote-ref-3)
5. Las bases se encuentran divididas en dos muestras, una de entrenamiento (*train*) y otra de prueba (*test*), de cada una se contó con una base de datos a nivel de personas y otra a nivel hogar. Dado que la estimación que nos interesa es pobreza del hogar, se exponen las características en esta agregación. [↑](#endnote-ref-4)