**Problem Set No. 2 – Predicting Poverty**

# Introducción:

Por medio de este documento, se presentan los resultados del estudio de la Base de datos de Medición de Pobreza Monetaria y Desigualdad de 2018 para el territorio Colombiano, a través de diferentes modelos de clasificación y regresión. Estos modelos, fueron debidamente cargados a la plataforma de Kaggel en miras de competir con los demás frente a aquel con mejor ajuste.

## La predicción de la pobreza como un asunto relevante para la política pública:

En el contexto económico, político y social, la generación de programas, políticas y proyectos que busquen enfrentar la pobreza, es uno de los principales objetivos de la política pública a nivel nacional, regional e internacional[[1]](#footnote-2). En el ámbito internacional, 149 miembros de las naciones unidas ratificaron, a inicios del milenio, un compromiso multidimensional tendiente a combatir la pobreza, a partir de la adopción de esquemas gubernamentales y de cooperación comunes a nivel mundial, considerando como finalidad la erradicación del fenómeno[[2]](#footnote-3).

No obstante, la definición de un estándar objetivo donde se pueda determinar cuando una persona o un hogar se encuentra en una situación de pobreza, puede variar según los contextos propios de cada país[[3]](#footnote-4). De allí, que la definición y cuantificación de como una persona u hogar es pobre, permite generar un marco de información mucho más claro para que las instituciones estatales promuevan las acciones necesarias para enfrentarla.

En el caso colombiano la definición de la pobreza se fundamenta en un estándar a partir de la línea de pobreza, como indicador desarrollado por el DANE. Así, un hogar, cuyos ingresos se encuentre por debajo de este valor, será para estos efectos un hogar “pobre”. Cabe aclarar que este indicador, esto es, el de cada hogar, se calcula tomando los ingresos que recibe el hogar, y dividiéndolo entre el número de individuos que lo habita, para efectos de calcular el ingreso monetario de cada individuo. En lo que respecta al cálculo de la línea de pobreza, este indicador es definido por el DANE como la cantidad de dinero que necesita una persona para solventar sus necesidades básicas[[4]](#footnote-5). A esto se le conoce como el indicador de pobreza monetaria, al estar relacionado con los ingresos el hogar y la persona.

Adicionalmente, en Colombia existe otro indicador ampliamente utilizado, relativa a la pobreza multidimensional, no obstante, en este trabajo nos concentraremos en la pobreza monetaria[[5]](#footnote-6).

## Resumen de los resultados y estructura del documento:

El estudio se fundamentó en la construcción de un modelo predictivo frente a la pobreza de los hogares a partir de la siguiente ecuación:

Donde I es la función de ingresos del hogar y, reiterando que esta debe ser menor a la línea de pobreza descrita para los hogares en igual condición.

Del análisis realizado se obtuvieron los siguientes resultados:

[xx]

Para llegar a lo anterior, y considerando lo expuesto, el documento se dividirá en los siguientes acápites, a saber: (i). Presentación del enlace de acceso al repositorio en GitHUB; (ii). Proceso de limpieza de datos y estadísticas descripticas; (iii). Modelo de clasificación; (iv). Modelo de regresiones; (v). Conclusiones.

# Enlace a repositorio de GitHub:

El enlace de Github donde podrá encontrarse el repositorio con las respuestas del taller es el siguiente:

|  |
| --- |
| **Enlace al repositorio en GitHub** |
| <https://github.com/Carlosvergara1995/Problem_Set_2-_Predicting_Poverty.git> |

# Descripción de los datos:

A continuación, se presenta el proceso de limpieza de datos y las correspondientes estadísticas descriptivas:

## Origen de los datos:

Para efectos de este taller, se utilizaron los datos de la Gran Encuesta Integrada de Hogares – GEIH 2018, y dimensión particular para hogares y personas en relación con la pobreza monetaria y desigualdad, ambas creadas por el DANE, según reporte de 2019. Las bases de datos estudiadas se dividían en dos grupos: (i). Bases de datos de testeo (testing) y (ii). Bases de datos de entrenamiento (training), (en adelante los “Grupos”). Cada grupo se componía en dos bases de datos correspondientes a personas y hogares.

## Proceso de limpieza de datos:

Inicialmente, y dado el tamaño de las bases, y las limitaciones de carga del aplicativo GitHub, se modifican las bases de formato cvs a rds.

Considerando las instrucciones del taller, se procedió a tomar la base de hogares y personas de cada uno de los grupos, con el propósito de que la información concerniente a personas complementara la relativa a hogares. Para estos efectos, y considerando que las variables en las bases de entrenamiento y testeo deben ser iguales en cuanto a la existencia de variables independientes, se verificaron las variables de cada una de las bases, se identificó a las variables categóricas y posteriormente se modificaron las variables necesarias para unir la base de datos y de personas en cada Grupo.

Se advierte que, en el caso de la base de testeo, no se contaba con la variable “Pobre”, que para efectos de este ejercicio es la variable dependiente. Igualmente, se pudo identificar que existían muchas variables que no existían en la base de testeo, pero si en las de entrenamiento, por lo que se procedió a sustraerlas.

De todo este proceso surgen dos bases unificadas, correspondientes a hogares con el complemento de los datos de las bases de personas, tanto para el grupo de testeo, como para el grupo de entrenamiento.

Luego, y una vez unidas las bases, se identificaron los *missing values* en cada una, para remplazarlos.

Una vez realizado lo anterior, y dada la existencia de variables categóricas, estas se modifican para convertirlas en variables dummy, para efectos de poder ejecutar los análisis de clasificación y regresión. Con esto, se guardan las bases de datos finales: dvf\_test de testeo y dvf\_trainign= para entrenamiento.

## Estadísticas descriptivas:

A continuación, se presenta de manera breve las estadísticas de las bases de datos finales:

# Modelo de clasificación:

# Modelo de regresiones:

# Conclusión:

1. Janeth Patricia Muñiz Eraso. *La pobreza y las políticas públicas: del referencial global al sectorial.* Cuadernos del CENDES, Vol. 32, No. 88, Caracas, Venezuela. (2015). Disponible en: <http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1012-25082015000100006#:~:text=A%20nivel%20mundial%20existe%20un,metas%20de%20las%20pol%C3%ADticas%20p%C3%BAblicas>. [↑](#footnote-ref-2)
2. Naciones Unidas, Asamblea General. Resolución No. A/RES/55/2 de 13 de septiembre de 2000, *Declaración del Milenio.* Al respecto, indicó la Asamblea General en los conciderandos 11 y siguientes del mencionado acto: “11. *No escatimaremos esfuerzos para liberar a nuestros semejantes, hombres, mujeres y niños, de las condiciones abyectas y deshumanizadoras de la pobreza extrema, a la que en la actualidad están sometidos más de 1.000 millones de seres humanos. Estamos empeñados en hacer realidad para todos ellos el derecho al desarrollo y a poner a toda la especie humana al abrigo de la necesidad.*

   *12. Resolvemos, en consecuencia, crear en los planos nacional y mundial un entorno propicio al desarrollo y a la eliminación de la pobreza.*

   *13. El logro de esos objetivos depende, entre otras cosas, de la buena gestión de los asuntos públicos en cada país. Depende también de la buena gestión de los asuntos públicos en el plano internacional y de la transparencia de los sistemas financieros, monetarios y comerciales. Propugnamos un sistema comercial y financiero multilateral abierto, equitativo, basado en normas, previsible y no discriminatorio*.

   [...]” [↑](#footnote-ref-3)
3. Erika Yurany Villaizon Castro. *Evolución Metodológica de la Medición de la Pobreza en Colombia.* Rev. Inclusión & Desarrollo, No. 4, (2015). [↑](#footnote-ref-4)
4. Departamento Nacional de Planeación, Dirección de Desarrollo Social. *Pobreza monetaria y pobreza multidimensional: Análisis 2008-2018.* (2019). Disponible en: <https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Social/Documento%20de%20An%C3%A1lisis%20de%20las%20Cifras%20de%20Pobreza%202018.pdf> [↑](#footnote-ref-5)
5. Según el Departamento Nacional de Planeación: *El Índice de Pobreza Multidimensional (IPM), desarrollado por el Oxford Poverty & Human Development Initiative (OPHI), es un indicador que refleja el grado de privación de las personas en un conjunto de dimensiones. La medida permite determinar la naturaleza de la privación (de acuerdo con las dimensiones seleccionadas) y la intensidad de la misma EL IPM es la combinación del porcentaje de personas consideradas pobres, y de la proporción de dimensiones en las cuales los hogares son, en promedio, pobres)*. Ver: Departamento Nacional de Planeación. *Índice de Pobreza Multidimensional (IMP-Colombia). 1997-2008 y meta del PND para 2014.* (s.f.). Disponible en: <https://colaboracion.dnp.gov.co/CDT/Estudios%20Econmicos/%C3%8Dndice%20de%20Pobreza%20Multidimensional%20(IPM-Colombia)%201997-2008.pdf> [↑](#footnote-ref-6)