

Problem Set 2: Predicting Poverty

I. Introducción

Este estudio se basa en la competencia “Pover-T Tests: Predicting Poverty” realizada por el Banco Mundial, la cual tuvo como objetivo la predicción de la pobreza en Colombia. Este análisis partió de la hipótesis de que los altos costos y el tiempo generan dificultades para la medición de esta, lo que complejiza obtener resultados apropiados. Sin embargo, al desarrollar mejores modelos, se pueden realizar encuestas con menos preguntas y más específicas que midan de manera rápida y económica la efectividad de las nuevas políticas e intervenciones. Por esta razón, es importante tener mayor precisión para orientar estas intervenciones e iterar las políticas, maximizando su impacto y su rentabilidad.

Diseñar una política pública es fundamental para la construcción de una sociedad justa y en la que todos tengan oportunidades. Por ello, es importante la ejecución acertada de modelos que puedan predecir correctamente la población objetivo y que dicha política pueda ser aplicada para realizar intervenciones a las familias adecuadas.

En este documento se presentan dos modelos de predicción de la pobreza en los hogares colombianos para conocer con exactitud cuáles hogares son pobres y cuáles no, con el fin de que la política pueda ser aplicada a quienes se encuentran realmente en condición de pobreza, y así, puedan recibir las ayudas del Gobierno, planteadas en la política, evitando casos en donde algunos hogares no sean identificados como pobres y con ello, no reciban los subsidios respectivos. El primer modelo se ejecutará por medio de clasificación de hogares pobres, y el segundo, se realizará por medio de una regresión, en donde se toman los ingresos y se comparan con la línea de pobreza para definir si son pobres o no.

Así, a lo largo de este *Problem Set*, se realizará un modelo de predicción de cuáles hogares son los más pobres, basado en características individuales con datos adquiridos por la Gran Encuesta Integrada de Hogares (GEIH) de 2018 del Departamento Administrativo Nacional de Estadística (DANE), para conocer información sociodemográfica de la población y señalar casos de fraude que podrían conducir a la reducción de la brecha en el país e identificar a las familias vulnerables que podrían necesitar mayor asistencia por parte del Gobierno. Para ello, se aplicarán dos metodologías: Clasificación y Regresión. Además, se identificarán parámetros como ROC, falsos positivos, falsos negativos y demás elementos para predecir de la manera más acertada los hogares que serían objeto de análisis de las políticas relacionadas con este problema.

Nota:

La base de datos usada, al igual que el script de R y el presente documento están disponibles en el repositorio de GitHub en el siguiente enlace: <https://github.com/Yilap/Taller2-BDML.git>

Contexto

Para entender esta problemática es importante considerar que la pobreza es una condición socioeconómica de vulnerabilidad y carencia de recursos con efectos negativos en la salud, las relaciones sociales, la productividad y la acumulación de capital humano (Universidad ICESI, 2020). Según el Banco Mundial (2018), para el año 2015, el porcentaje de personas que vivían con menos de 1,95 dólares al día alcanzó 10% de la población mundial. Para el caso de Colombia, y teniendo en cuenta las estadísticas del Departamento Administrativo Nacional de Estadística-

DANE (2019), el porcentaje de personas que se ubicaron por debajo de la línea de pobreza alcanzó el 27% en el 2018.

Para comprender este fenómeno, es importante entender que existe diferencia entre la tasa de incidencia de pobreza entre grandes y pequeñas ciudades, tal es el caso de ciudades pequeñas como Quibdó y Riohacha, donde la tasa de incidencia superó el 47% en 2018, mientras que en Bogotá y Medellín estuvo por debajo del 14% (DANE, 2019). Por ello, el análisis de los factores asociados a la dinámica de la pobreza desde una perspectiva macro y micro a nivel de ciudades, resulta de gran interés y relevancia en el plano de la política pública nacional y local (Universidad ICESI, 2020).

Entre 2018 y 2022 según el DANE, se ha generado un fuerte incremento de la pobreza monetaria en Colombia, pues a diciembre de 2021 se registraron 19.621.000 personas que vivían con menos de 11.801 pesos al día, equivalentes a un 39,3 % de la población y, por otro lado, se registraron 6.111.000 personas con menos de 5.730 pesos al día, equivalentes al 12,2 % (Universidad Nacional de Colombia, 2022). También se evidencia que el 31 % de las personas se encuentran en vulnerabilidad monetaria, es decir, que viven con un ingreso diario que oscila entre 11.801 y 23.017 pesos al día. De acuerdo con lo anterior, alrededor de 35 millones de colombianos (70,3 % de la población) afrontan esta situación, siendo el 52% mujeres y 48% hombres (Universidad Nacional de Colombia, 2022).

Dado lo anterior, esta situación demuestra que el mercado laboral en Colombia ha experimentado fuertes incrementos llegando a cifras de dos dígitos, pues para el año 2021 el desempleo se ubicó en 13,8% y en el 2022 en 11,2%, reduciéndose en 2,6 puntos porcentuales, donde el 10,6 % se representó por los hombres y de 18,1 % por las mujeres (DANE, 2019). Este panorama indica que la pobreza es un fenómeno social que debe ser atendido por el Gobierno nacional a través de intervenciones eficientes que permitan la toma de decisiones basadas en evidencia.

II. Datos

a. Descripción de las fuentes de datos

Para el desarrollo de este *Problem Set* se utilizarán los datos de la Gran Encuesta Integrada de Hogares (GEIH) de 2018 del Departamento Administrativo Nacional de Estadística (DANE). Esta encuesta contiene información sobre las condiciones de empleo de las personas (si trabajan, en qué trabajan, cuánto ganan, fuentes de ingresos, si tienen seguridad social en salud o si están buscando empleo), adicional a las características generales de la población como sexo, edad, estado civil y nivel educativo (DANE, 2018). La GEIH consolida información no solo a nivel nacional sino a nivel regional, departamental, cabecera y ciudades capitales.

La GEIH es una fuente de información pertinente para el análisis de la pobreza en Colombia, ya que, al condensar información sociodemográfica sobre los individuos, los niveles de ingresos laborales y no laborales y la estructura de la fuerza de trabajo en el territorio nacional (tasas de ocupación, desempleo e informalidad laboral), permite identificar las características más relevantes que influyen sobre la renta personal. Usando estos datos como insumo, es posible predecir diferencias significativas entre los ingresos de los hogares y poder determinar si se encuentran en situación de pobreza o no.

a. Análisis descriptivo de los datos (estadísticas descriptivas)

Este trabajo tiene como propósito construir modelos predictivos de pobreza a nivel de hogar tomando como referencia los datos de la Gran Encuesta Integrada de Hogares - GEIH año 2018. Estos datos se encuentran segmentados en dos bases de datos, por un lado, se tiene la base *training* que proporciona información de la variable pobreza e ingreso, y por el otro lado, se tiene la base de datos *testing* que no proporciona estas variables y, por lo tanto, se realiza la predicción de los hogares pobres sobre esta última considerando el mejor modelo entrenado.

Para la base de datos *training* se realizaron los siguientes ajustes: i). se colapsa la base de datos de personas para obtener información agrupada por hogar de las variables objeto de análisis y se consolida con la base de datos de hogares; ii). se crean las siguientes variables para facilitar el análisis de la base de datos: PorcentajeOcupados, ViveEnCabecera, JefeMujer, PersonaPorCuarto, TipoVivienda, RegimenSalud, EducaciónPromedio, AntigüedadTrabajo, TipoDeTrabajo, Pobre, Lp, Ingpcug; iii). se procede a realizar ajuste de los missing values de cada variable, registrando el cero en estos valores para las siguientes variables. Sin embargo, se observó que la única variable que presenta missing values es en Régimen de Salud, lo cual no afecta el poder estadístico.

Para la base de testing, se replicaron los mismos ajustes mencionados anteriormente, pues el análisis se hizo bajo las mismas variables para ambas bases de datos. De esta manera, se eligieron alrededor de 21 variables, las cuales se consideran que tienen relevancia para el estudio de predicción de la pobreza, es decir, estas variables pueden servir para identificar si un hogar es pobre o no.

Tabla 1. Estadísticas descriptivas

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ViveEnCabecera*	1	164960	1.91	0.29	2.0	2.00	0.00	1.0	2.0	1.0	-2.79	5.77	0.00
JefeMujer*	2	164960	1.42	0.49	1.0	1.40	0.00	1.0	2.0	1.0	0.33	-1.89	0.00
PersonaPorCuarto	3	164960	1.73	0.83	1.5	1.59	0.74	0.2	16.0	15.8	2.30	10.47	0.00
TipoVivienda*	4	164960	2.46	1.26	3.0	2.39	1.48	1.0	6.0	5.0	0.12	-1.16	0.00
RegimenSalud	5	164960	1.54	0.50	2.0	1.55	0.00	0.0	2.0	2.0	-0.15	-1.98	0.00
TipoDeTrabajo*	6	164960	3.11	1.92	2.0	3.02	1.48	1.0	10.0	9.0	0.44	-0.82	0.00
Pobre	7	164960	0.20	0.40	0.0	0.13	0.00	0.0	1.0	1.0	1.50	0.25	0.00
PorcentajeOcupados	8	164960	0.50	0.32	0.5	0.50	0.25	0.0	1.0	1.0	0.17	-0.89	0.00
AntigüedadTrabajo	9	164960	78.39	119.48	24.0	51.67	35.58	0.0	948.0	948.0	2.02	4.07	0.29
Lp	10	164960	271522.31	33656.89	279944.5	280449.07	7410.67	167222.5	303816.7	136594.2	-2.56	5.07	82.87
IngresoPerCapita	11	164960	870639.26	1244349.74	543568.5	645894.35	434812.98	0.0	88833333.3	88833333.3	10.21	302.27	3063.75
EducaciónPromedio	12	164960	9.34	3.69	9.0	9.25	3.46	0.0	16.0	16.0	0.08	-0.42	0.01

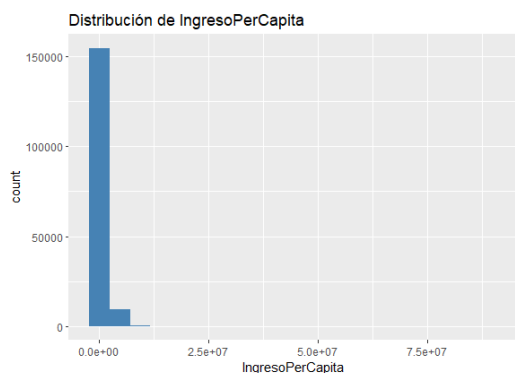
Fuente: Elaboración propia con R

A partir de la tabla anterior se puede inferir que para el análisis se tuvieron en cuenta 164.960 observaciones (hogares) de la base training. Se presentan 12 variables y se obtiene la siguiente información relevante: el ingreso per cápita promedio por hogar es de 870.639, los años promedio de educación son 9 y máximo 16 años.

En las siguientes gráficas se puede evidenciar que el 20% corresponde a hogares clasificados como pobres y el 80% restante a hogares no pobres, el 57% de los hogares en promedio sus individuos no cuentan con trabajo formal y que el 42% de los hogares tienen como jefe de hogar a las mujeres. Por otra parte, se muestra que los hogares pobres disminuyen a medida que aumenta el nivel de educación, siendo más representativa la brecha de ingresos, pues las personas tienen mas oportunidad laboralmente.

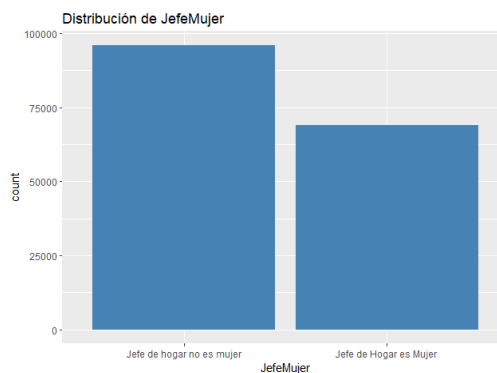
De acuerdo con lo anterior, se considera que existe una relación entre ser pobre, tener más hijos, trabajar menos horas a la semana, no contar con trabajo formal y no contar con vivienda propia, esto a su vez se estaría viendo reflejado en menos ingresos promedio de los hogares. Finalmente, comparando estas estadísticas generales de la base *training* respecto a la base *testing*, se concluye que no existen diferencias significativas en la media de las variables presentadas, lo que quiere decir que ambas bases son muy similares.

Gráfica No. 1 Ingresos per cápita de hogar



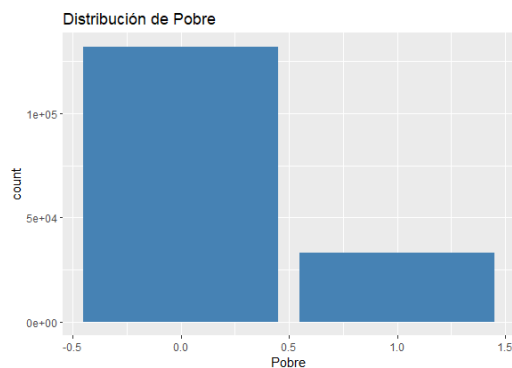
Fuente: Elaboración propia con R Studio

Gráfica No. 2 Distribución jefe



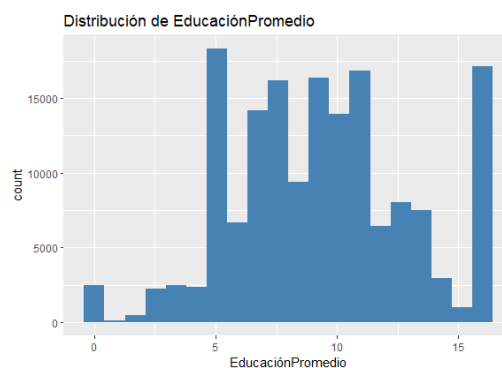
Fuente: Elaboración propia con R

Gráfica No. 3 Distribución pobre Educación promedio



Fuente: Elaboración propia con R Studio con R Studio

Gráfica No. 4 Distribución



Fuente: Elaboración propia

b. Modelo de Clasificación

En esta sección se realiza la predicción de hogares pobres y no pobres mediante modelos de clasificación binarios. Se evaluaron tres modelos con diferentes conjuntos de variables predictoras, estos fueron:

Modelo	Especificación
M1	$Pobre = \beta_0 + \beta_1 PorcentajeOcupados + \beta_2 JefeMujer + \beta_3 PersonaPorCuarto + \beta_4 TipoVivienda + \beta_5 ViveEnCabecera + \beta_6 TipoDeTrabajo + \beta_7 RegimenSalud + \beta_8 EducaciónPromedio + \beta_9 AntigüedadTrabajo$
M2	$Pobre = \beta_0 + \beta_1 PorcentajeOcupados + \beta_2 JefeMujer + \beta_3 PersonaPorCuarto + \beta_4 TipoVivienda + \beta_5 TipoDeTrabajo + \beta_6 RegimenSalud + \beta_7 EducaciónPromedio$
M3	$Pobre = \beta_0 + \beta_1 PorcentajeOcupados + \beta_2 JefeMujer + \beta_3 PersonaPorCuarto + \beta_4 TipoVivienda + \beta_5 RegimenSalud + \beta_6 EducaciónPromedio$

Para cada una de las especificaciones se corrieron múltiples modelos de aprendizaje automático utilizando diferentes algoritmos y técnicas de regularización, específicamente Logit, Lasso (priorizando la sensibilidad y el ROC), Lasso con down sampling y up sampling, y Elastic Net.

Debido al desbalance de clases, donde cerca del 80% de los hogares son clasificados como pobres y solo el 20% no lo son, se ha dividido la muestra de entrenamiento en tres partes: i) training, que representa el 70% de las observaciones de la base de datos principal y se usó para estimar los modelos; ii) evaluación, la cual fue útil para desarrollar técnicas de post procesamiento, evaluando el cutoff óptimo de los modelos y, por último, iii) testing, para predecir la pobreza utilizando los modelos estimados y con el objetivo de seleccionar el mejor modelo.

En la tabla siguiente (ordenada de mayor a menor *accuracy* por efectos del concurso de *Kaggle*, el cual prioriza esta métrica), se presentan los resultados de los diferentes modelos y métricas de evaluación. Dado que es una muestra desbalanceada, el *accuracy* no es la única métrica relevante por considerar. En este caso, un buen modelo es el que logra un equilibrio entre la sensibilidad (capacidad de detectar verdaderos positivos) y la especificidad (capacidad de detectar verdaderos negativos), así como una alta precisión (proporción de predicciones correctas). Asimismo, otras métricas importantes a considerar son el área bajo la curva ROC (AUC-ROC) y el coeficiente kappa.

Tabla 2. Métricas

Modelo	Alpha	Lambda	ROC	Sens	Spec	Accuracy	Kappa
lasso_roc1	0	0,014	0,87736	0,95591	0,42121	0,84887	0,44415
lasso_roc2	0	0,014	0,87742	0,95575	0,42077	0,84865	0,44340
elasticnet1	0,1	0,000	0,87810	0,94047	0,47869	0,84803	0,46900
logit1	NA	NA	0,87804	0,93974	0,48120	0,84795	0,46992
logit2	NA	NA	0,87812	0,93941	0,48185	0,84781	0,46991
elasticnet2	0,1	0,029	0,87818	0,93999	0,47908	0,84772	0,46845
logit3	NA	NA	0,87070	0,94108	0,45836	0,84444	0,45120
elasticnet3	0,1	0,029	0,87065	0,94164	0,45516	0,84425	0,44930
lasso_roc3	0	0,014	0,87735	0,95580	0,42070	0,84210	0,41230
lasso2	0	1,023	0,86930	0,99613	0,09673	0,81608	0,13916
lasso1	0	1,023	0,86896	0,99596	0,09724	0,81604	0,13959

lasso3	0	1,023	0,86560	0,99647	0,08474	0,81395	0,12263
lasso_downsample2	0	0,024	0,87601	0,77718	0,81131	0,79424	0,58848
lasso_upsample1	0	0,024	0,87722	0,77744	0,81080	0,79412	0,58823
lasso_upsample2	0	0,026	0,87705	0,77716	0,81044	0,79380	0,58760
lasso_downsample1	0	0,032	0,87555	0,77744	0,80984	0,79364	0,58727
lasso_upsample3	0	0,038	0,87037	0,77289	0,80438	0,78863	0,57727
lasso_downsample3	0	0,026	0,86923	0,77272	0,80434	0,78853	0,57706

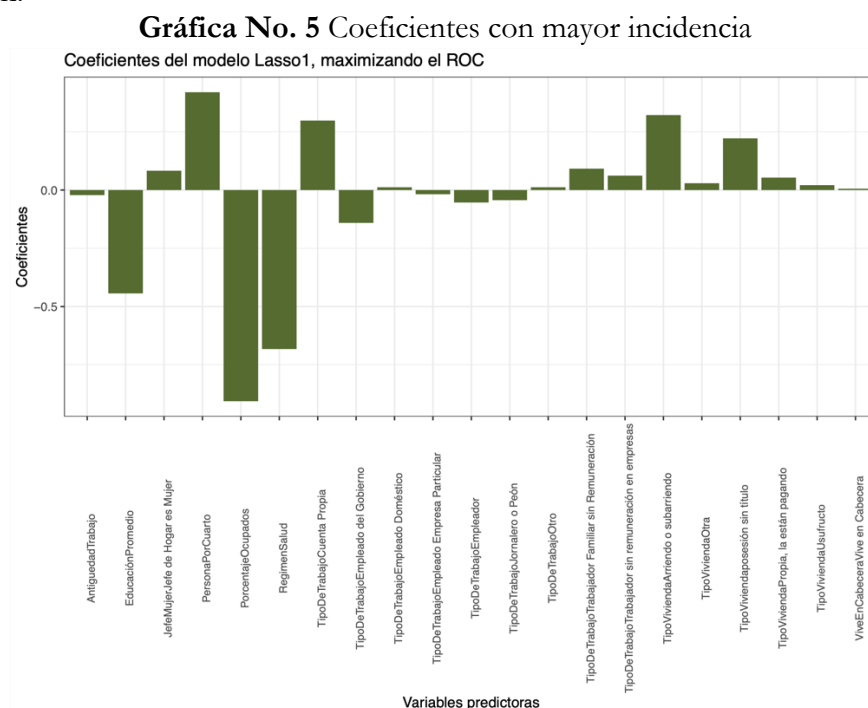
En este caso, algunos modelos principales que parecen tener un buen desempeño son:

- logitM1 y logitM2: Estos dos modelos tienen resultados muy similares en todas las métricas evaluadas, y su ROC es la más alta entre los modelos presentados.
- elasticnet1: Este modelo tiene una alta sensibilidad y una precisión razonable, y su ROC es también alta.
- lasso_roc1 y lasso_roc2: Estos modelos tienen la misma precisión y su ROC es la más alta entre los modelos de regularización Lasso.

La elección del mejor modelo en cada caso depende del contexto específico del problema. En este caso, se analiza el modelo “lasso_roc1” y se encuentra que las variables más importantes para la predicción de la pobreza son:

- **Reducen la probabilidad de que el hogar sea pobre:** % de personas ocupadas en el hogar, régimen de salud del jefe del hogar y años de educación promedio.
- **Aumentan la probabilidad de que el hogar sea pobre:** número de personas por cuarto, vivienda en arriendo o subarriendo, al igual que posesión sin título, y ser trabajador por cuenta propia.

Estos resultados coinciden con la teoría económica. Lo anterior se muestra en la gráfica a continuación:



c. *Modelo de Regresión*

En el modelo de regresión se evaluaron 4 tipos de enfoques regresión lineal con upsampling, regresión lineal con downsampling, árbol de decisión y bosques aleatorios, estos se aplicaron sobre el modelo lineal que se muestra a continuación:

$$\begin{aligned} \text{IngresoPerCapita}_h &= \beta_0 + \beta_1 \text{PorcentajeOcupados}_h + \beta_2 \text{ViveEnCabecera}_h + \beta_3 \text{JefeMujer}_h + \beta_4 \text{PersonaPorCuarto}_h \\ &+ \beta_5 \text{TipoVivienda}_h + \beta_6 \text{RegimenSalud}_h + \beta_7 \text{EducaciónPromedio}_h + \beta_8 \text{AntigüedadTrabajo}_h \\ &+ \beta_9 \text{TipoDeTrabajo}_h \end{aligned}$$

Donde h son los hogares, finalmente, para determinar si un hogar es pobre o no, se compara el IngresoPerCapita del hogar con la variable "línea de pobreza", si el ingreso per cápita es menor a la línea de pobreza, entonces el hogar es pobre, y viceversa.

Dado que la base de datos de entrenamiento presenta un desbalance (80% de los hogares no es pobre y el 20 % si es pobre), en el enfoque de regresión lineal en donde hay susceptibilidad de predecir erróneamente por desbalance se aplicó dos estrategias de resampling, por oversampling (en donde aumentamos el número de observaciones de la clase minoritaria) y undersampling (en donde eliminamos observaciones de la clase dominante). Los resultados de los modelos se muestran a continuación.

Tabla 3. Métricas

Modelo	Accuracy	Sensitivity	Specificity
OLS-Oversampling	0,8351	0,5345	0,9103
OLS- Downsampling	0,8344	0,5322	0,9101
Árbol	0,8017	0,0000	1,0000
Random Forest	0,8509	0,3595	0,9740

En el preste problem_set nos enfocamos en los criterios de comparación entre entrenamiento y prueba como lo son el Accuracy el cual mide la cantidad de aciertos de los Verdaderos positivos y verdaderos negativos; Sensitivity el cual mide la cantidad de Verdaderos positivos correctamente predichos y por último el Specificity el cual mide la cantidad de aciertos de los verdaderos negativos. Como se puede apreciar en la tabla, en accuracy los modelos OLS tienen resultados similares siendo el más preciso OLS-Oversampling, sin embargo, el más preciso de los 4 modelos es el de Bosques aleatorios o Random Forest; respecto a sensibilidad, son más acertados los modelos OLS y en especificidad el más poderoso es el árbol de decisión seguido de árboles aleatorios.

En los resultados resalta la ineficiencia del árbol de decisión a la hora de predecir, en comparación de sus pares, sin embargo, esto es de esperarse dada las limitaciones de usar un único árbol para la predicción. Si bien los modelos OLS tienen buen desempeño, dado que los bosque aleatorios tienen mayor accuracy, decidimos usar este modelo como el predictivo para la competencia de Kaggle, esto en gran medida porque este es el parámetro de interés para el concurso.

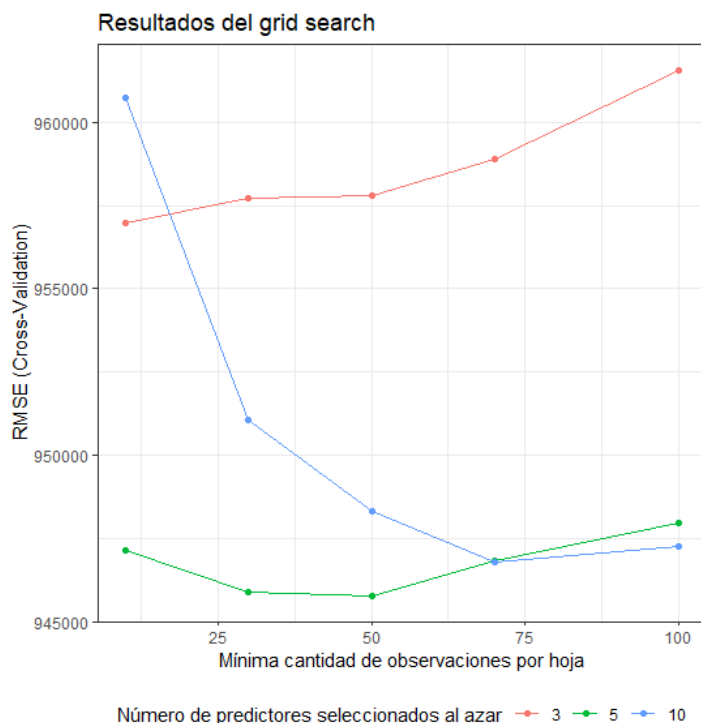
III. Modelo Final

Para el modelo final decidimos tomar random forest dado que el accuracy reportado entre sus pares es superior, adicionalmente, respecto a los modelos implementados por clasificación tiene la gran ventaja de manejar eficientemente las variables no lineales como por ejemplo el ingreso per cápita, la educación promedio, antigüedad en el trabajo entre otras.

Para el entrenamiento del modelo se utilizó una partición de la data original del 70% para entrenamiento y el 30% restante para testing. Para el tuning del modelo se usó el tuning óptimo del enfoque hallado

en función de la minimización del RMSE obtenido mediante Cross-Validation, según la optimización del modelo, los resultados óptimos se hallan para arboles con 5 predictores usados aleatoriamente, adicionalmente, la cantidad mínima de observaciones óptima es de aproximadamente 30, a continuación, se muestra el performance de la optimización para el grid search:

Gráfica No. 6 Resultados del grid search



Comparando nuestro modelo random forest con los mejores de clasificación según accuracy (Lasso Roc1 y Lasso Roc2) encontramos que los de clasificación si bien no superar a Random Forest en el parámetro de interés, son mucho mejores con la sensibilidad, finalmente comparando a Random Forest con los OLS de predicción por regresión del ingreso, encontramos que también los OLS son ligeramente mejores en sensibilidad sin embargo Random forest los supera en especificidad. Por último, cabe resaltar que la gran ventaja de random forest sobre al ser un modelo no paramétrico no requiere cumplir supuestos de normalidad o cualquier otro supuesto sobre la distribución de los datos.

Respecto a las variable usadas, podemos notar que el accuracy de los distintos modelos usados es similar, mostrando que no hay grandes diferencias en su base predictiva que finalmente son las variables usadas y creadas para el ejercicio, la causa del bueno performace de los modelos es atribuible al poder predictivo de la pobreza y el ingreso que tienen las mimas.

Finalmente, para cerrar la descripción del modelo cabe mencionar que otra de las grandes ventajas de este modelo es que en bueno prediciendo inclusive con muestras desbalanceadas por lo que hacer resampling no fue necesario.

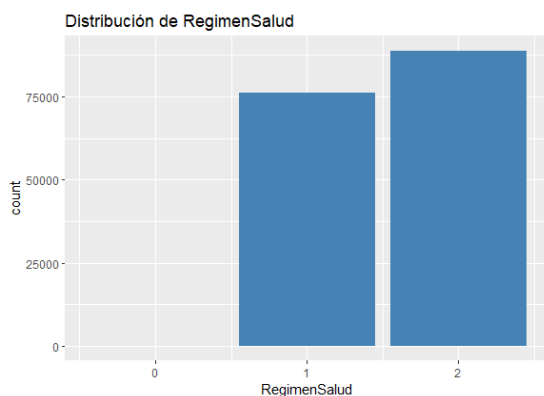
IV. Conclusión y recomendaciones

En este documento se abordaron múltiples modelos para predecir si un hogar es pobre o no en Colombia. El modelo seleccionado como óptimo para este ejercicio fue un Random Forest para predecir el ingreso y así determinar si este está o no por debajo de la línea de pobreza.

Consideramos que el uso de modelos predictivos robustos es un ejercicio fundamental para el gobierno colombiano, con la intención no solo de identificar a la población con más necesidades, sino también de asignar de manera eficiente los recursos disponibles para la inversión en este tipo de programas que mejoran el bienestar social y que para cumplir su objetivo deben estar bien focalizados.

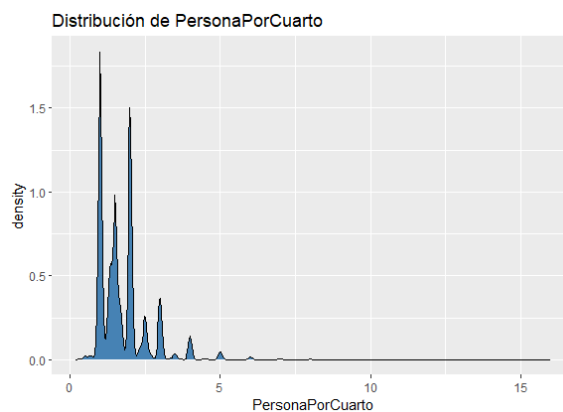
V. Anexos

Gráfica No.5 Régimen de salud



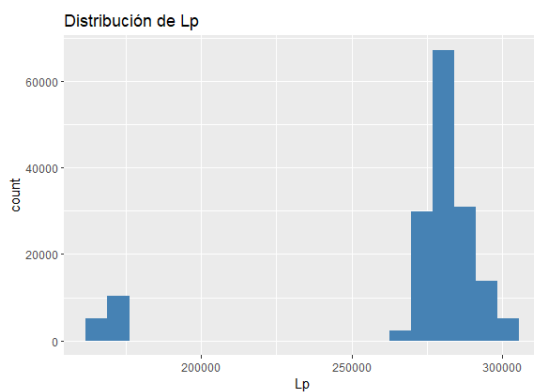
Fuente: Elaboración propia con R Studio

Gráfica No. 6 Distribución personas por cuarto



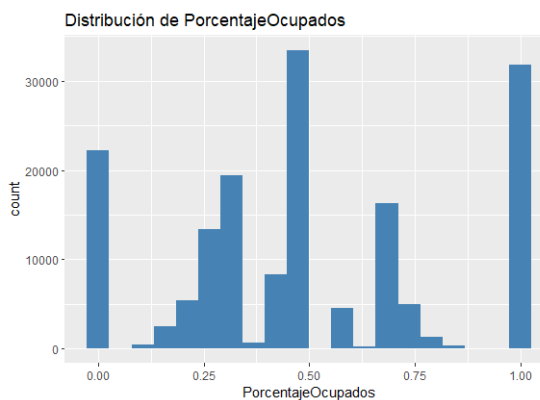
Fuente: Elaboración propia con R Studio

Gráfica No. 7 Distribución de Lp



Fuente: Elaboración propia con R Studio

Gráfica No. 8 Distribución porcentaje ocupados

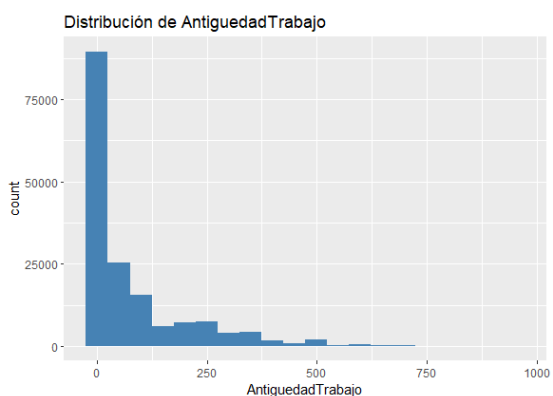


Fuente: Elaboración propia con R Studio

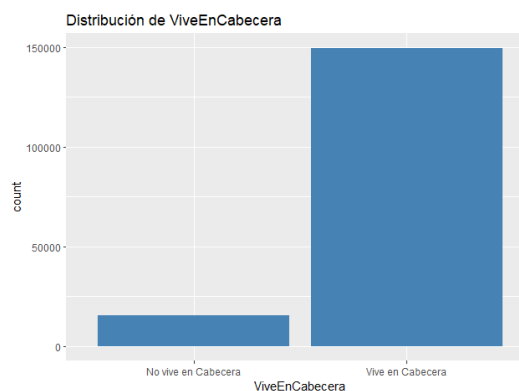
Gráfica No. 9 Distribución antigüedad trabajo

Gráfica No. 10 Distribución vice en cabecera

Presentado por Yilmer Palacios, Betina Cortés, Lida Jimena Rivera, Nelson Fabián López



Fuente: Elaboración propia con R Studio



Fuente: Elaboración propia con R Studio

VI. Bibliografía

Banco Mundial. (2018). La pobreza y la prosperidad compartida 2018: Armando el rompecabezas de la pobreza, panorama general. Banco Mundial, Washington, DC. Recuperado el 17 de marzo de 2018 de: <https://www.bancomundial.org/es/news/>

CEPAL. (2019). Panorama Social de América Latina. Santiago, Chile.

DANE., (2019). Boletín Técnico Pobreza Monetaria en Colombia Año 2018. Bogotá D.C., mayo3 de 2019. Tomado de: https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2018/bt_pobreza_monetaria_18.pdf

DNP. (2016). Conpes 3877. Declaración de importancia estratégica del sistema de identificación de potenciales beneficiarios (Sisbén IV). Bogotá.

Fedesarrollo. (2021). Descifrar el futuro. La economía colombiana en los próximos 10 años. Bogotá: Pinguin Ramdon House.

Universidad Nacional de Colombia., (2022). Población pobre en Colombia pierde 10 % de su ingreso por la inflación. Recuperado de: <https://periodico.unal.edu.co/articulos/poblacion-pobre-en-colombia-pierde-10-de-su-ingreso-por-la-inflacion/>