Juan Sebastián Durán Durán

Pablo Cerezo Lesmes

**Problem Set 2. Problem Set 2: Predicting Poverty “Wars of nations are fought to change maps. But wars of poverty are fought to map change” M. Ali MECA 4107**

¿Cómo podemos elaborar herramientas que efectivamente clasifiquen a las personas de acuerdo con parámetros establecidos? ¿Qué utilidad podemos obtener de los modelos de predicción?

En política pública, por ejemplo, tener un buen mecanismo de clasificación puede hacer la diferencia entre proveer efectivamente los recursos de un Estado o de manera análoga -e indeseable- privar de un programa a personas en condición de vulnerabilidad. En Colombia, por ejemplo, un mecanismo de clasificación de personas es el SISBEN. Por medio de esta herramienta, las personas son incluidas en categorías para focalizar la inversión. En ese sentido, el problema de una mala clasificación radica en que, las personas ricas podrían acceder a programas sociales, mientras que, aquellas personas de menores ingresos quedarían excluidas de estos programas.

En concreto, este ejercicio busca realizar una buena predicción a partir de la clasificación de personas pobres y no pobres. Los datos que se utilizan surgen a partir de la encuesta de “Medición de la Pobreza Monetaria y Desigualdad” para el año 2018 elaborada por el DANE. El reto específico de este ejercicio consiste en que, a partir de la información suministrada, se desarrollen modelos de predicción tales qué se minimice la probabilidad de hacer una clasificación errada – falsos positivos y falsos negativos. En principio se suministraron 4 bases de datos:

|  |  |  |
| --- | --- | --- |
| **Base de Datos:** | **Número de observaciones** | **Número de variables** |
| test\_hogares | 66168 | 16 |
| test\_personas | 219169 | 63 |
| train\_hogares | 164960 | 23 |
| train\_personas | 543584 | 135 |

En referencia a la tabla anterior, las bases de datos de entrenamiento y testeo están a nivel de personas y de hogares. Así mismo, vale la pena resaltar que, estos datos no tienen un mismo tamaño de observaciones ni tampoco contienen las mismas variables. Lo anterior implica que, en primer lugar, se debe explorar la información contenida en estas bases. En segundo lugar, se debe definir intuitivamente que variables son las necesarias para un buen modelo los modelos que se diseñaran y, finalmente, unir en una misma base, la información de hogares y personas. Una vez finalizada la exploración, y limpieza de base de datos, resultamos con la información suficiente para entrenar, evaluar y testear los modelos predictivos con respecto a la pobreza y el ingreso.

Conclusiones principales:

Lo anterior, se logra a partir de entrenar modelos, evaluarlos y finalmente testearlos.

para que una persona acceda a un programa y obtenga algún tipo de beneficio u un elemento que se utiliza elaborar mecanismos de clasificación de las personas permiteA partir de modelos de entrenamiento, evaluación y testeo