

CONSUMO DE AGUA POTABLE EN EL CASCO URBANO DE TOLEDO NORTE DE SANTANDER SEGUNDO SEMESTRE 2020

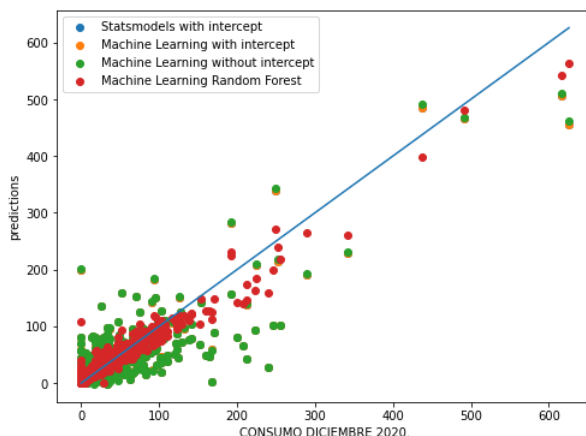
Cueto Rodriguez Rafael David, Maestría en estadística aplicada y ciencia de datos de la UTB, Ciencias básicas. E-mail: calidadsoluciones@gmail.com – rcueto@utb.edu.co

Descripción del problema: Se quiere predecir el consumo de agua potable de los hogares en el mes de diciembre del 2020 en el casco urbano de Toledo, teniendo en cuenta el consumo de agua de los meses de junio, agosto, septiembre, octubre y noviembre del 2020.

Descripción de los datos: Los datos se obtuvieron del sitio web <https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Consumo-de-agua-potable-en-el-casco-urbano-segundo/3dma-ihgg>, donde se tienen dos variables cualitativas y siete numéricas. En esta investigación la variable a predecir es el “Consumo de agua potable del mes de diciembre de 2020” y se utilizarán los consumos de agua obtenidos desde el mes de junio a noviembre del 2020 en el casco urbano del municipio de Toledo ubicado dentro del departamento de Norte de Santander, Colombia.

Metodología: En esta investigación se aplica machine learning con regresión lineal múltiple y machine learning con Random Forest.

Resultado: Una vez aplicado los modelos de regresión lineal se puede observar que el mejor modelo es el *machine learning con Random forest*, el cual muestra un SCORE más cercano a uno con valor de 0.948 y un RMS más cercano a cero con valor de 9.388, lo anteriormente mencionado se evidencia con el grafico y los resultados del modelo.



| | | |
|-----------|----|--------------------|
| score_SM | is | 0.7046544663581416 |
| score_LR | is | 0.7046543922418571 |
| score_LR2 | is | 0.7043221739026936 |
| score_RF | is | 0.9482841274973423 |
| RMS2_SM | is | 22.435719640322286 |
| RMS2_LR | is | 22.435722455418308 |
| RMS2_LR2 | is | 22.44833727583196 |
| RMS2_RF | is | 9.38829502196906 |