# Aplicación de ML Sobre Información Georeferenciada. Predicción de Frecuencia de Cortes Eléctricos

•••

Informe de Mentoría 2020
Diplomatura en Ciencias de Datos FaMAF-UNC

Ariel Rubio
Andrés Ruderman
Sacha Smrekar

Mentor: Ramiro Caro

# Objetivo

Determinar las principales causas que explican la frecuencia de corte del suministro eléctrico en una región determinada de Brasil.

# Road Map

Aislar la variable de referencia que describe los cortes

Limpiar y agrupar los dataset

Generar nuevas variables

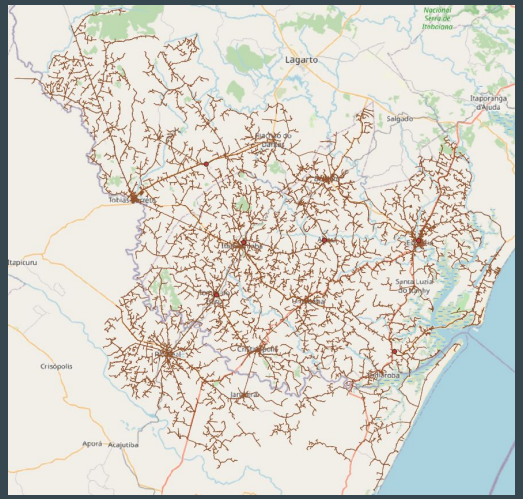
Utilizar herramientas de Machine learning para aislar las principales causas que expliquen los cortes

# Dataset

El dataset fue provisto por la Agencia Nacional de Energía Eléctrica de Brasil. Consta de datos georeferenciados entre los que destacan:

- Unidad Consumidora de Baja Tensión (UCBT)
- Segmento del Sistema de Distribución de Baja Tensión (SSDMT)
- Unidad Transformadora de Distribución (UNTRD)
- Unidad Transformadora de Subestación (UNTRS)
- Segmento Conductor (SEGCON)

Entre otras.

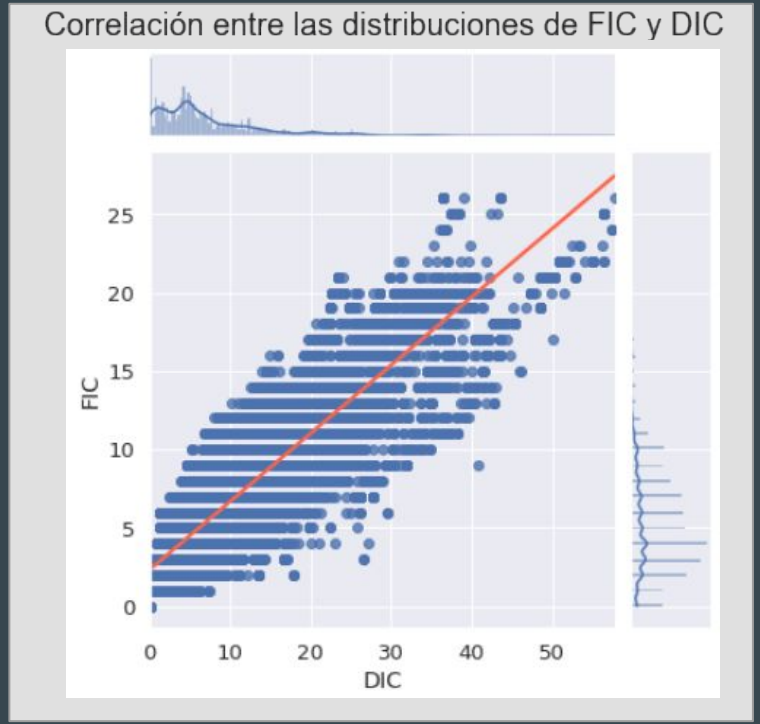



Red eléctrica de Itabaianinha, Sergipe, Brazil

# Variable Objetivo

Los factores que inciden en los cortes de energía se pueden estudiar utilizando las  variables FIC (Frecuencia de interrupción del servicio) y DIC (Duración de la interrupción del servicio).

Ambas variables están correlacionadas, por lo que elegimos trabajar con la variable FIC.



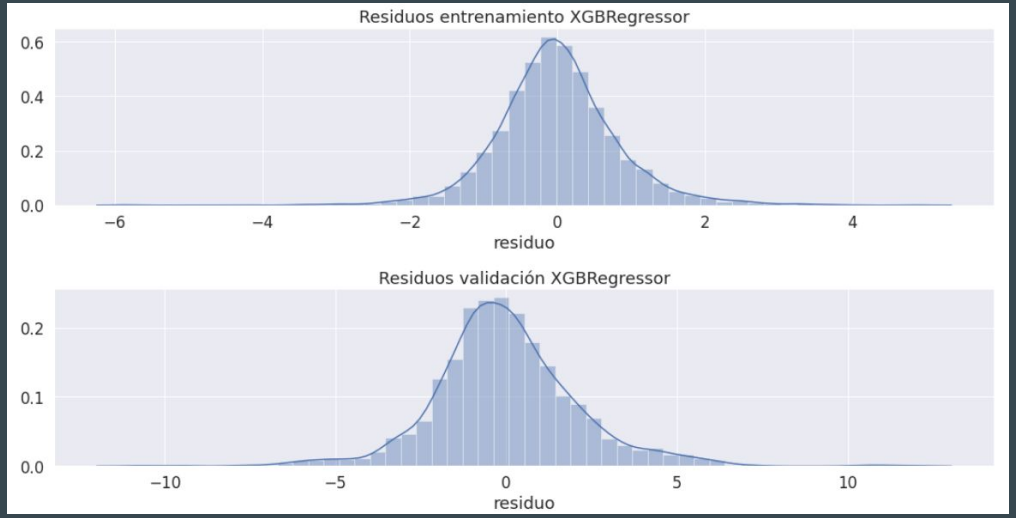
# Agrupamiento del dataset y generación de variables

- Seleccionamos para trabajar el dataset de unidades de transformación de distribución (UNTRD).
- Variables estadísticas agrupando la información de los consumidores de baja tensión (UCBT).
- Variables asociadas a la unidad de transformación de subestación (UNTRS).
- Variables georeferenciadas: distancia a carreteras, distancia a ciudades, densidad de UCBT cercanos, etc.
- Variables asociadas a las propiedades de la red (análisis de grafos), así como variables que tuvieran en cuenta la distancia medida a través de la red.

# Primeros Modelos

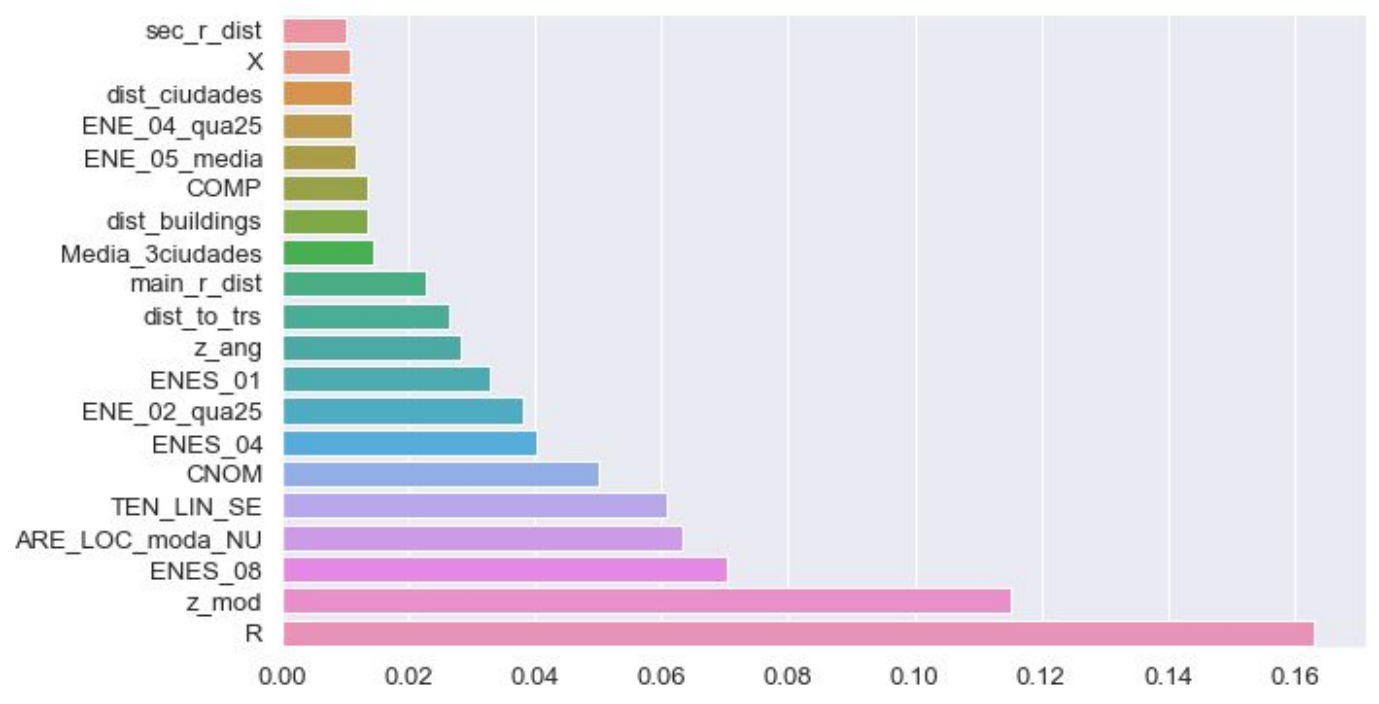
Como la variable objetivo no está distribuida en clases son los algoritmos de regresión los que se ajustan a nuestro problema.

| Modelo | MAE |
|---|---|
| Regresión Lineal | 2.73 |
| Regresión Lineal con regularización L2 | 2.74 |
| Tree Boosting (XGBRegressor) | 1.54 |

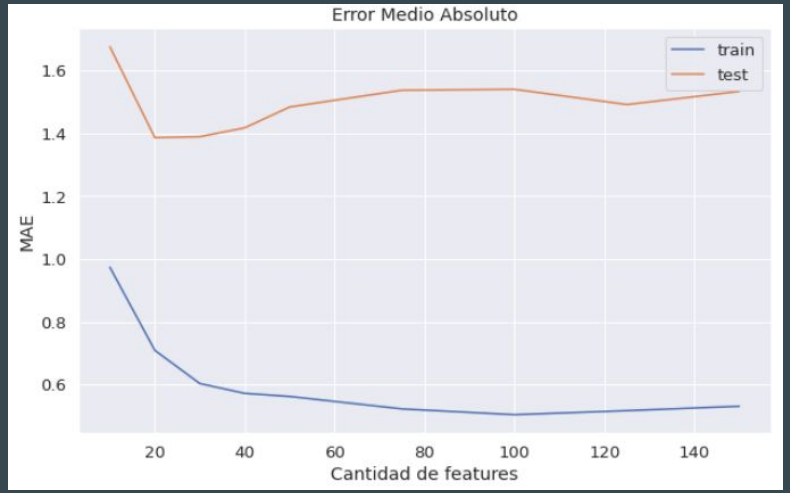



Seleccionamos el algoritmo XGBRegressor para generar los modelos de aprendizaje

# Variables más relevantes (XGBRegressor)



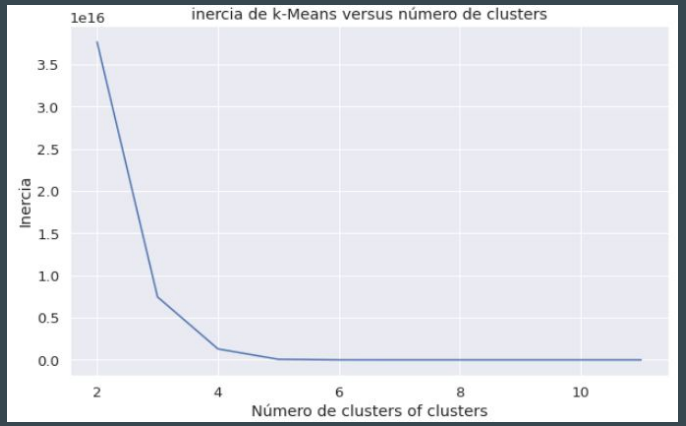
# Optimización del modelo

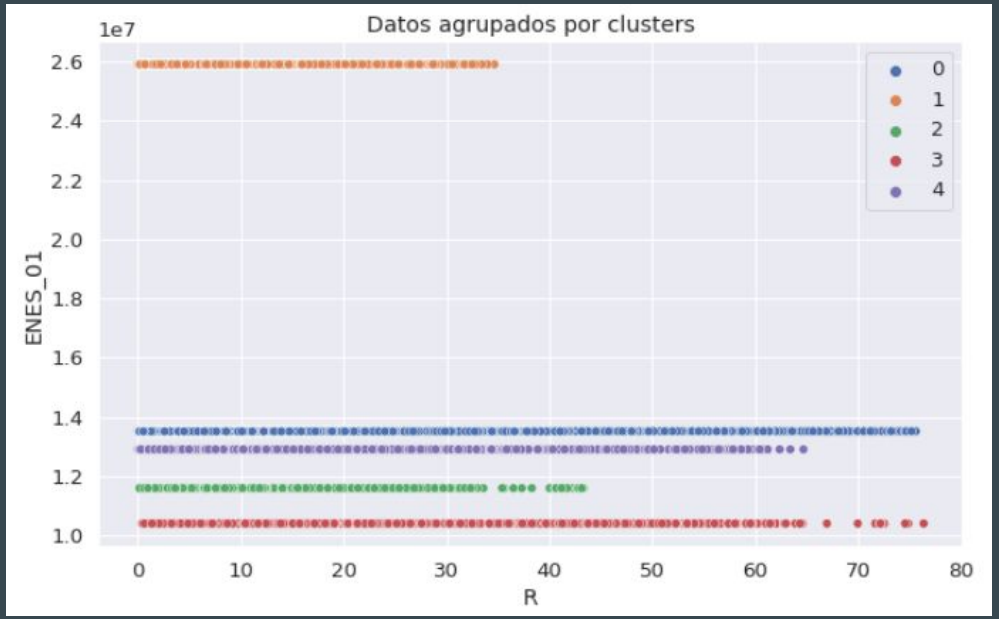



Seleccionamos las 20 mejores variables y optimizamos los hiperparámetros del modelo. El MAE fue: 1.34

| param_colsample_bytree | param_max_depth | param_min_child_weight | param_min_split_loss | param_subsample | mean_test_score | std_test_score | rank_test_score |
|---|---|---|---|---|---|---|---|
| 0.8 | 14 | 4 | 0.75 | 1 | 0.800117 | 0.020184 | 4 |
| 0.8 | 14 | 6 | 0.75 | 1 | 0.799469 | 0.019984 | 5 |
| 0.9 | 8 | 1 | 0.01 | 1 | 0.801118 | 0.019859 | 1 |
| 0.9 | 14 | 5 | 0.1 | 1 | 0.800279 | 0.020285 | 3 |
| 1 | 14 | 2 | 0.5 | 1 | 0.800459 | 0.017373 | 2 |

# Clustering



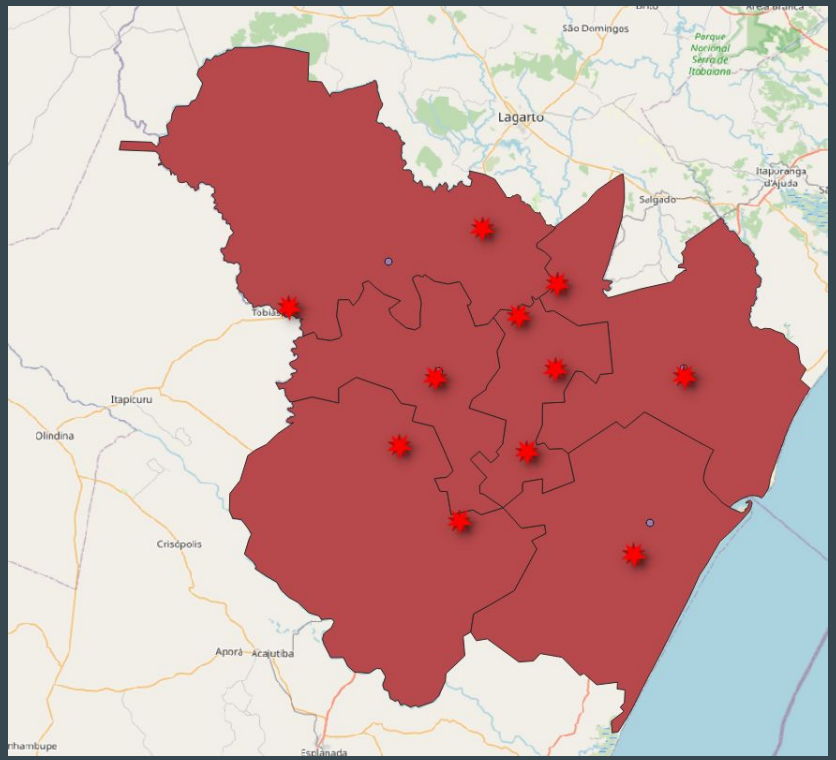

Utilizando técnicas de clustering redujimos el MAE a 1.27

# Dificultades halladas

Tuvimos problemas a la hora de identificar puntos importantes como ciudades, subestaciones distribuidoras o a línea costera utilizando los datos georeferenciados.

Este problema se resolvió utilizando QGIS para asignar estas ubicaciones a mano.

Como perspectiva a futuro nos gustaría automatizar esta
designación utilizando algún algoritmo.



# Conclusiones

1. Visualizamos la distribución de los datos, identificamos la variable objetivo FIC y determinamos que el modelo de regresión es el que mejor aplica a nuestro problema.
2. Generamos nuevas variables cruzando diferentes datasets e incorporamos los datos georeferenciados.
3. Determinamos que el  mejor algoritmo para modelar el problema es el XGBRegressor.
4. Mediante la selección de las variables más relevantes, el ajuste de hiperparámetros y el agrupamiento de datos mediante técnicas de clustering pudimos reducir el considerablemente el MAE.