

## Desarrollo

Se decide usar jupyter como editor de código. Decido cargar algunas librerías para cargar los datos, visualización de datos y modelos analíticos. Dan 3 datasets los cuales contienen información sobre cuentas bancarias las cuales hay que identificar si hay fraude o no en una transacción, Me dan un dataset ya entrenado con las variables que se tienen que analizar, por lo que empiezo con este. Decido ver cuales eran las variables son numéricas la mayoría de ellas son numéricas y solo hay 3 que contienen texto de 47 variables. Se decide ver que contienen las variables con texto y se convierten a valores numéricos, muchas de ellas tienen valores nulos por lo que se reemplazan por 0, estas no son transformadas en dummy, se les da un peso a cada una de las variables que contenga cada una de las columnas con texto.

Al ver que no había mucha correlación entre variables, toma la decisión de iniciar con el modelo, la visualización de datos no me iba a dar mucha información al respecto.

Al hacer el modelo tome todas las variables del dataset menos radicado y fraude, ya que, no iban a representar valor en el entrenamiento, se obtuvo más de un 80% de precisión lo que da un poco de seguridad al dar los valores que se quieren predecir. También uso el `cross_val_predict` para poder evaluar los residuos del modelo y diagnosticar su comportamiento. El error de testeo dio bajo entonces se puede asegurar que hay una buena relación entre las variables y es un buen modelo.

Sabiendo esto simplemente aplico los mismos conceptos para intervenir el dataset que se desea testear y con las mismas variables se obtienen los resultados del modelo que aprendió. Al final solo queda añadir estos valores al dataset `base_evaluada` para terminar con la prueba.