**Análisis predictivo**

**a. Determinación de la base de datos** Se utilizó un conjunto de datos de clientes de una entidad bancaria con 10,000 registros y 15 variables, de las cuales "Exited" representa la variable objetivo (si el cliente abandonó o no).

**b. Preprocesamiento y limpieza**

* Se codificaron variables categóricas como "Geography", "Gender" y "Card Type" mediante One-Hot Encoding.
* Se dividió la data en conjunto de entrenamiento (80%) y prueba (20%).

**c. Análisis descriptivo**

**d. Selección de variables** Se evaluó la importancia de cada variable en la predicción del abandono. Se realizaron dos escenarios:

* **Con "Complain" incluida**: Mejor capacidad predictiva.
* **Sin "Complain"**: Se analizó su influencia para evitar data leakage.

**e. Selección de Modelos** Se aplicaron tres modelos supervisados:

**1. Regresión Logística**

* Con "Complain": Accuracy 99.90%, recall clase 1 = 1.00
* Sin "Complain": Accuracy 85.95%, recall clase 1 = 0.51

**2. Random Forest**

* Con "Complain": Accuracy 99.90%, recall clase 1 = 1.00
* Sin "Complain": Accuracy 86.70%, recall clase 1 = 0.46

**3. XGBoost**

* Con "Complain": Accuracy 99.90%, recall clase 1 = 1.00
* Sin "Complain": Accuracy 85.65%, recall clase 1 = 0.51

**Conclusiones del Análisis Predictivo**

* Todos los modelos con la variable "Complain" alcanzaron una precisión superior al 99%, lo que demuestra su alto poder predictivo.
* Al eliminar "Complain", el desempeño bajó ligeramente, destacando la importancia de esta variable.