

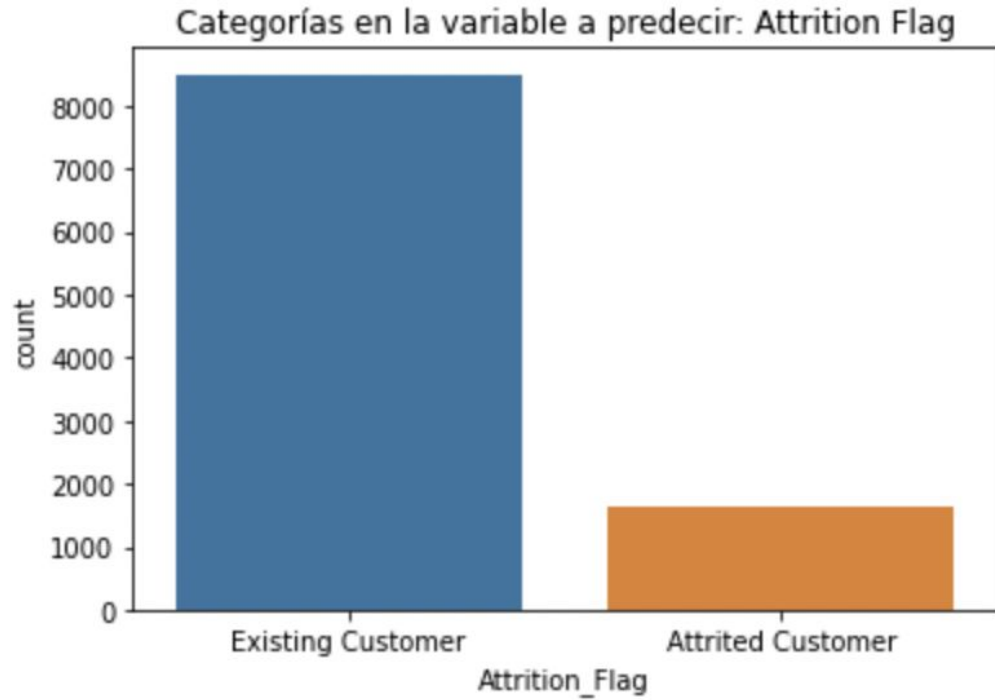
Proyecto Analítica Empresarial

Agustina Bentura, María Fabini,
Ximena Martinez, Belén Varela

Contexto del Problema

El Gerente del Banco está preocupado por la cantidad de clientes que abandonan los servicios de las tarjetas de crédito.

Éstos son aproximadamente un 16% de los clientes del banco.



Objetivos

1. Análisis Exploratorio para identificar el comportamiento de las variables y su relación con la variable principal (Attrited Customers)
2. Aplicar modelo con las variables seleccionadas (Correlación absoluta mayor a 0.01)
3. Medir la precisión del modelo



Dataset

El conjunto de datos utilizado consta de 10,000 clientes que mencionan su edad, salario, estado civil, límite de tarjeta de crédito, categoría de tarjeta de crédito, entre otros.

Problema del Dataset:

Solo tenemos un 16,07% de clientes que han cancelado. Por lo tanto, es un poco difícil entrenar nuestro modelo para predecir la rotación de clientes.

Además, tuvimos que predecir una variable categórica, por lo que usamos un modelo de regresión logística, en lugar de regresión lineal.



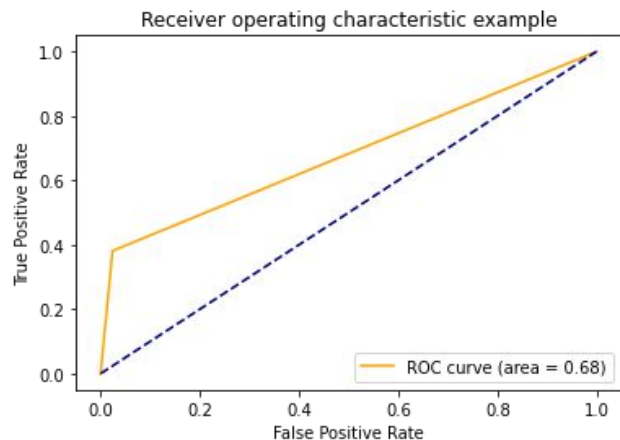
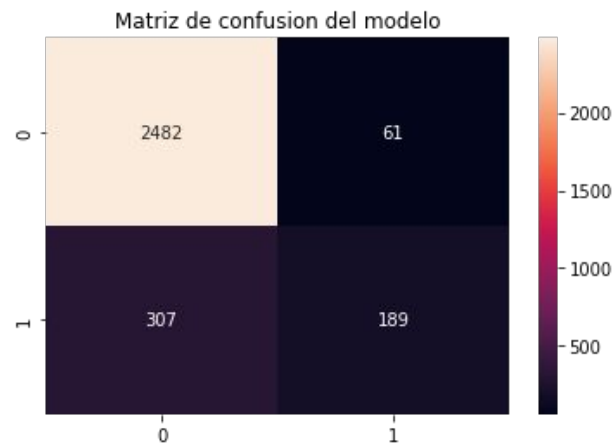
	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Education_Level	Marital_Status	Income_Category	Card_Category	Months_on_book	...	Credi
0	768805383	Existing Customer	45	M	3	High School	Married	\$60K - \$80K	Blue	39	...	
1	818770008	Existing Customer	49	F	5	Graduate	Single	Less than \$40K	Blue	44	...	
2	713982108	Existing Customer	51	M	3	Graduate	Married	\$80K - \$120K	Blue	36	...	
3	769911858	Existing Customer	40	F	4	High School	Unknown	Less than \$40K	Blue	34	...	
4	709106358	Existing Customer	40	M	3	Uneducated	Married	\$60K - \$80K	Blue	21	...	

5 rows x 23 columns

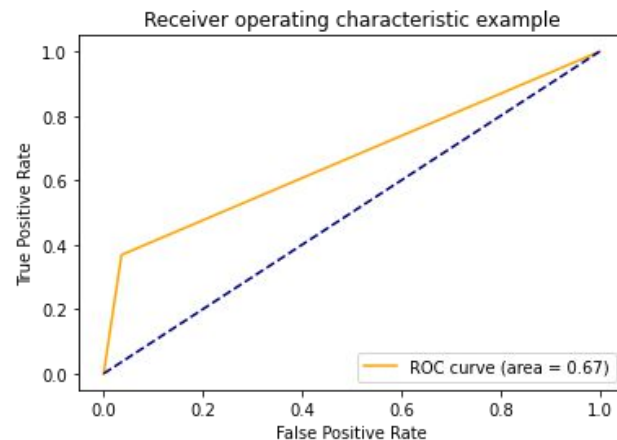
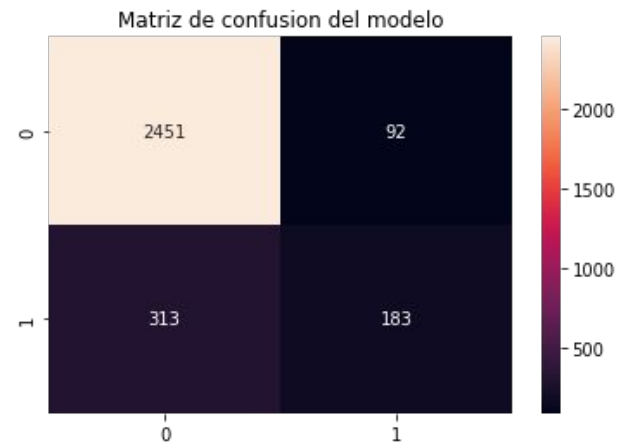
Análisis

1. Estudiamos la correlación entre las variables mediante un mapa de calor.
2. Seleccionamos las variables con una correlación absoluta mayor a 0.01, para realizar el entrenamiento del modelo y por otra parte utilizamos el data set entero para comparar.
3. Una vez realizada la regresión logística obtuvimos un porcentaje de accuracy de 86,67% para el dataset filtrado y un 87, 89% para el dataset completo.

Dataset Completo



Dataset Filtrado



Conclusión

Al comparar el modelo con ambos grupos, se puede ver como el dataset completo presenta una mayor precisión, lo que evidentemente es beneficioso para predecir la cantidad de clientes que abandonan los servicios del banco.

