



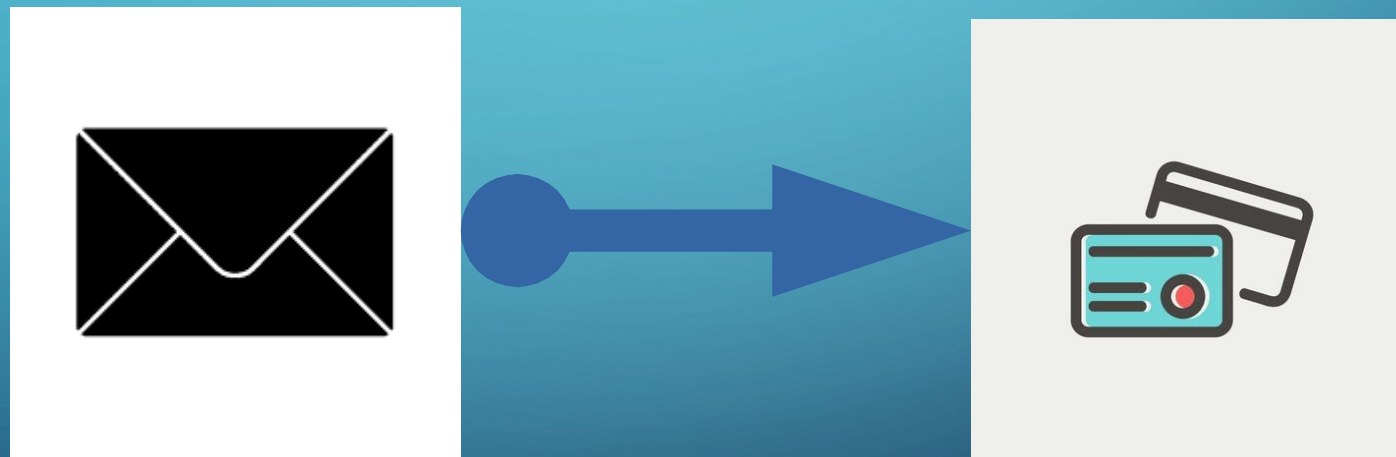
DIPLOMATURA DATA SCIENCE

GRUPO 2: VELAZQUEZ, GARCIA, PIZARRO

MODULO 2: APRENDIZAJE SUPERVISADO

OBJETIVO

Predecir qué clientes que a abrir un mail de marketing para ofrecer tarjetas de crédito



DATOS DISPONIBLES

Dataset
41 CAMPOS

- Fechas
- Info Plazos fijos
- Info Caja de ahorros
- Saldos en AR\$ U\$D
- Consumos débito y crédito



LIMPIEZA DE DATOS

En el proceso de limpieza se usó el propuesto
como plantilla

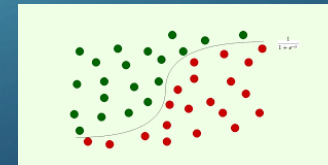
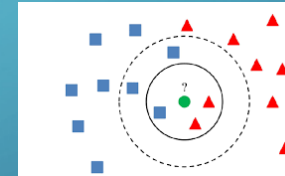
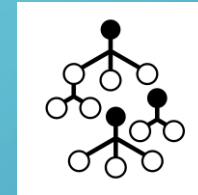
En una próxima entrega se mejorará



MODELADO

Se utilizaron 3 modelos:

- Random forest
- K Nearest Neighbors
- Logistic Regression



RESULTADOS PARCIALES

Logistic Regression

	precision	recall	f1-score	support
0.0	0.66	0.67	0.66	1905
1.0	0.38	0.37	0.37	1055
accuracy			0.56	2960
macro avg	0.52	0.52	0.52	2960
weighted avg	0.56	0.56	0.56	2960

K Nearest Neighbors

	precision	recall	f1-score	support
0.0	0.65	0.90	0.76	1905
1.0	0.42	0.13	0.20	1055
accuracy			0.63	2960
macro avg	0.54	0.51	0.48	2960
weighted avg	0.57	0.63	0.56	2960

Random forest

	precision	recall	f1-score	support
0.0	0.66	0.87	0.75	1905
1.0	0.46	0.19	0.27	1055
accuracy			0.63	2960
macro avg	0.56	0.53	0.51	2960
weighted avg	0.59	0.63	0.58	2960

Nada alentador, pero...

PASOS A SEGUIR

Mejorar EDA y limpieza de datos

Mejorar tratamiento de Outliers

Incluir Cross Validation

Generar nuevas variables desde categoricas

Probar diferentes escalados

FIN

MUCHAS GRACIAS !

