LOAN DEFAULTER PREDICTION

By: Maha Alatifi

13 January 2022

AGENDA

01
PROBLEM STATEMENT
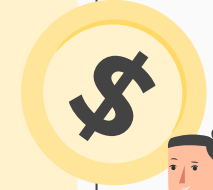
02
DATASET

03
MODELS & RESULTS

04
CONCLUSIONS

PROBLEM STATEMENT

The main idea of this project was to develop a framework by using classification models to predict the loan defaulter in banking to minimise the risk of losing money while lending to customers.

OVERVIEW

➢ Data preparation and defining the correlation between features.

➢ Training the dataset on 4 classifiers.

➢ Comparing between the models, in which the Random Forest showed the best score on F1.
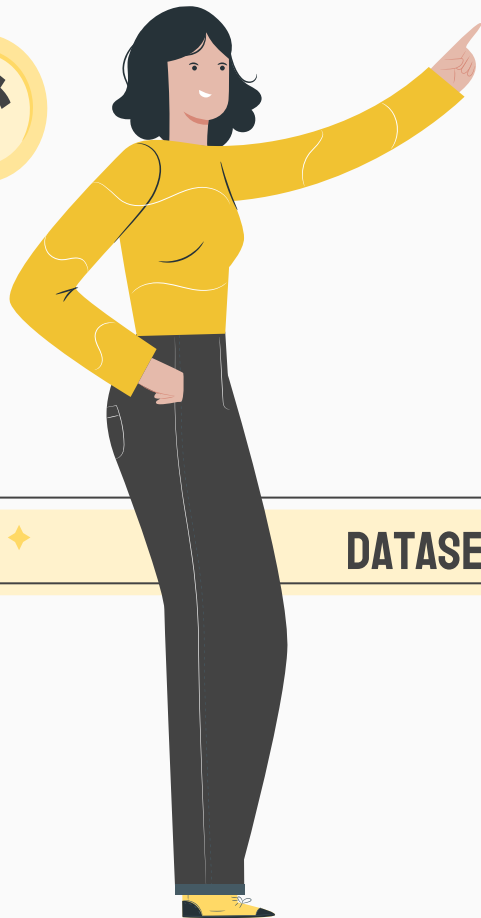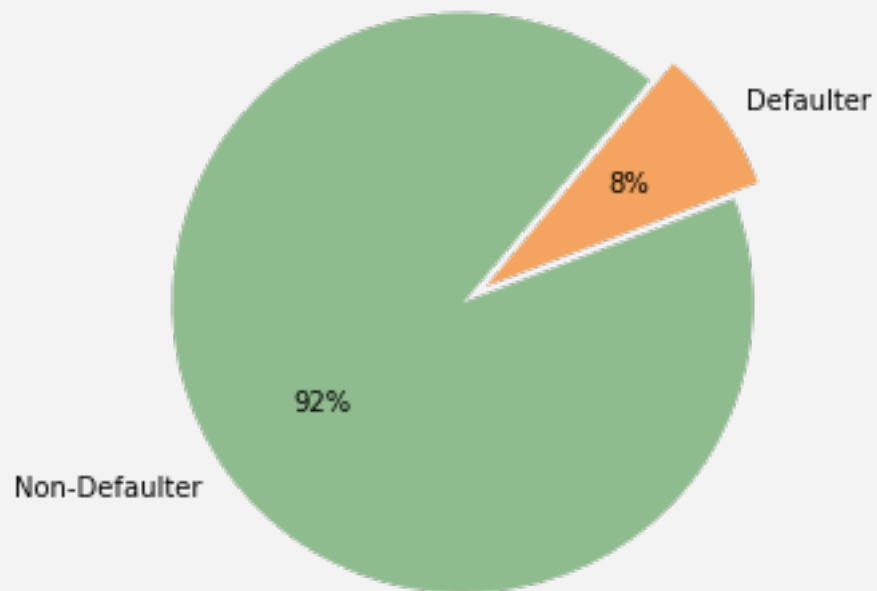
307,511

APPLICATIONS

122

FEATURES

DATASET SUMMERY

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
0	100002	1	Cash loans	M	N	Y	0	202500.0	406597.5	24700.5	351000.0
1	100003	0	Cash loans	F	N	N	0	270000.0	1293502.5	35698.5	1129500.0
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	135000.0	6750.0	135000.0
3	100006	0	Cash loans	F	N	Y	0	135000.0	312682.5	29686.5	297000.0
4	100007	0	Cash loans	M	N	Y	0	121500.0	513000.0	21865.5	513000.0
5	100008	0	Cash loans	M	N	Y	0	99000.0	490495.5	27517.5	454500.0
6	100009	0	Cash loans	F	Y	Y	1	171000.0	1560726.0	41301.0	1395000.0
7	100010	0	Cash loans	M	Y	Y	0	360000.0	1530000.0	42075.0	1530000.0
8	100011	0	Cash loans	F	N	Y	0	112500.0	1019610.0	33826.5	913500.0
9	100012	0	Revolving loans	M	N	Y	0	135000.0	405000.0	20250.0	405000.0

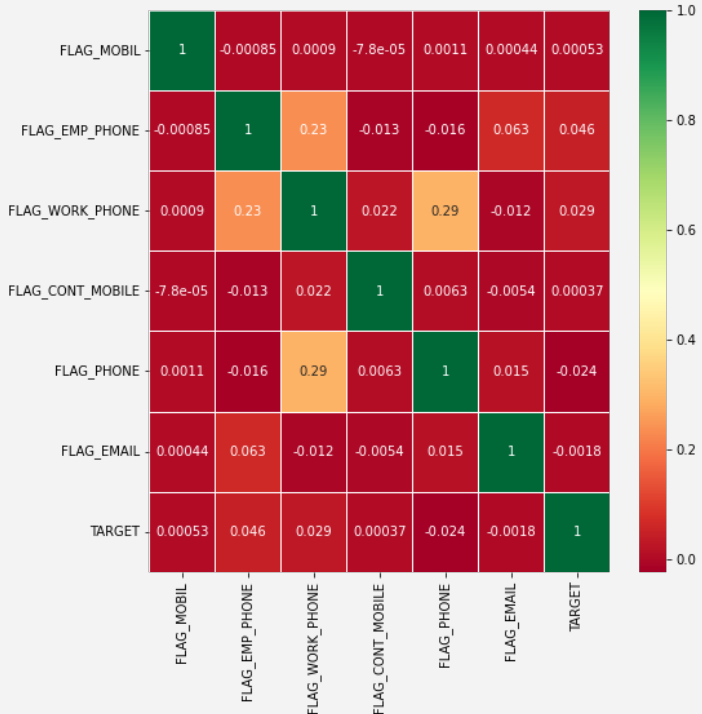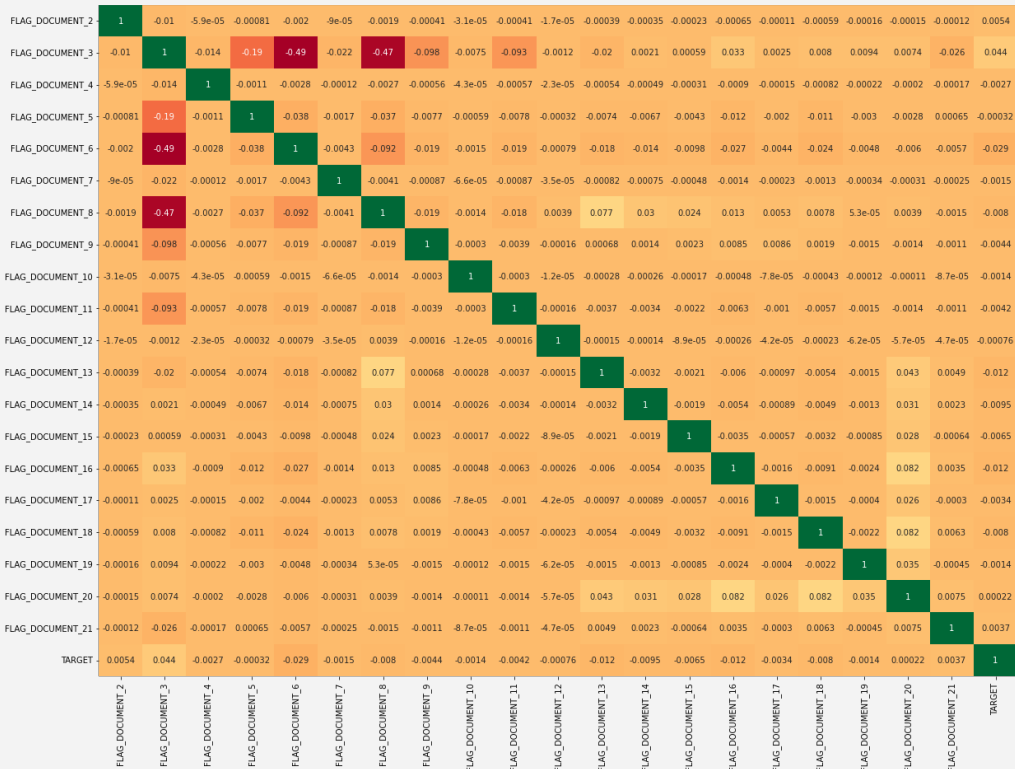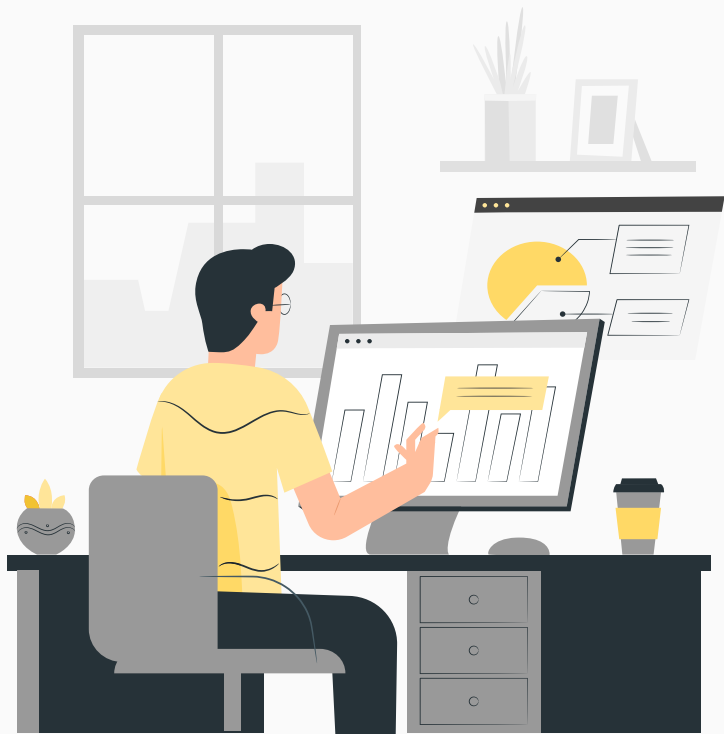# MODELS AND RESULTS

Percentage of Target variable

FEATURES CORRELATIONS

THE CLASSIFIERS MODELS

✓ Logistic regression (LR)

✓ K-nearest neighbours (K-NN)

✓ Decision Tree (DT)

✓ Random Forest (RF)

Confusion Matrix

Predicted Values

	Positive	Negative
Positive	TP Actual: Defaulter Predict: Defaulter	FN Actual: Defaulter Predict: Non_Defaulter
Negative	FP Actual: Non_Defaulter Predict: Defaulter	TN Actual: Non_Defaulter Predict: Non_Defaulter

Actual Values

RESULTS FOR 4 MODELS

Model	Function	Precision	Recall	Accuracy	F1
Logistic regression (LR)	LogisticRegression	0.87	0.54	0.91	0.64
K-nearest neighbours (K-NN)	KNeighborsClassifier	0.86	0.91	0.92	0.84
Decision Tree (DT)	DecisionTreeClassifier	0.88	0.61	0.84	0.70
Random Forest (RF)	RandomForestClassifier	0.85	0.92	0.92	0.88

THANKS !

Do you have any questions?