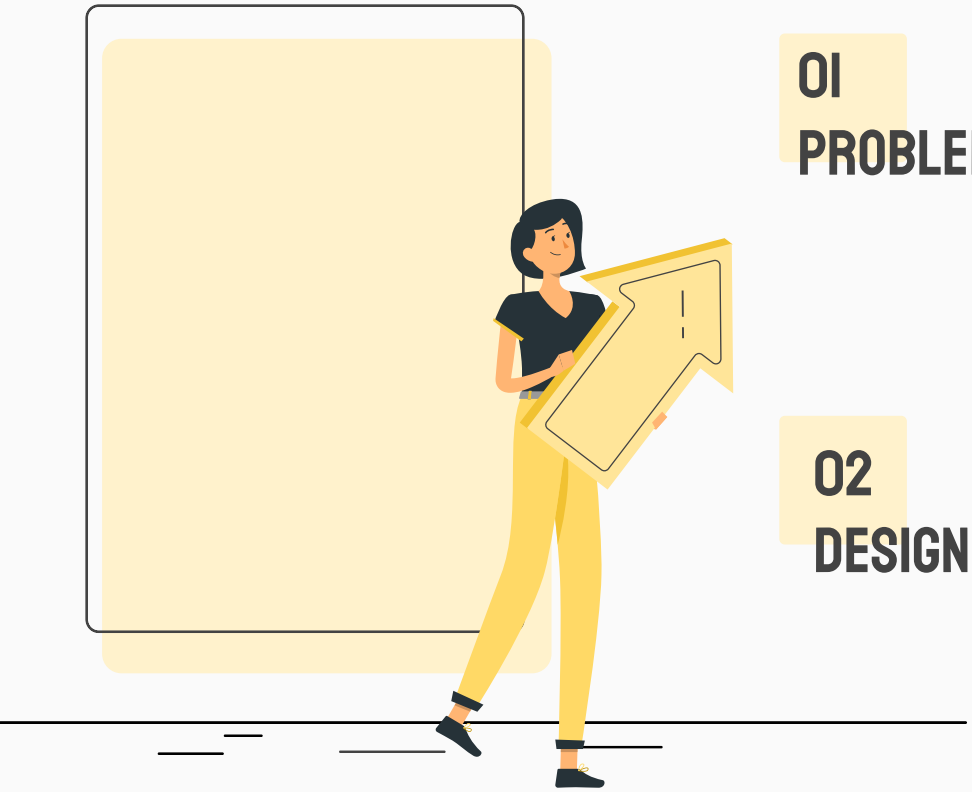
LOAN DEFAULTER PREDICTION

By: Maha Alatifi



01
PROBLEM STATEMENT

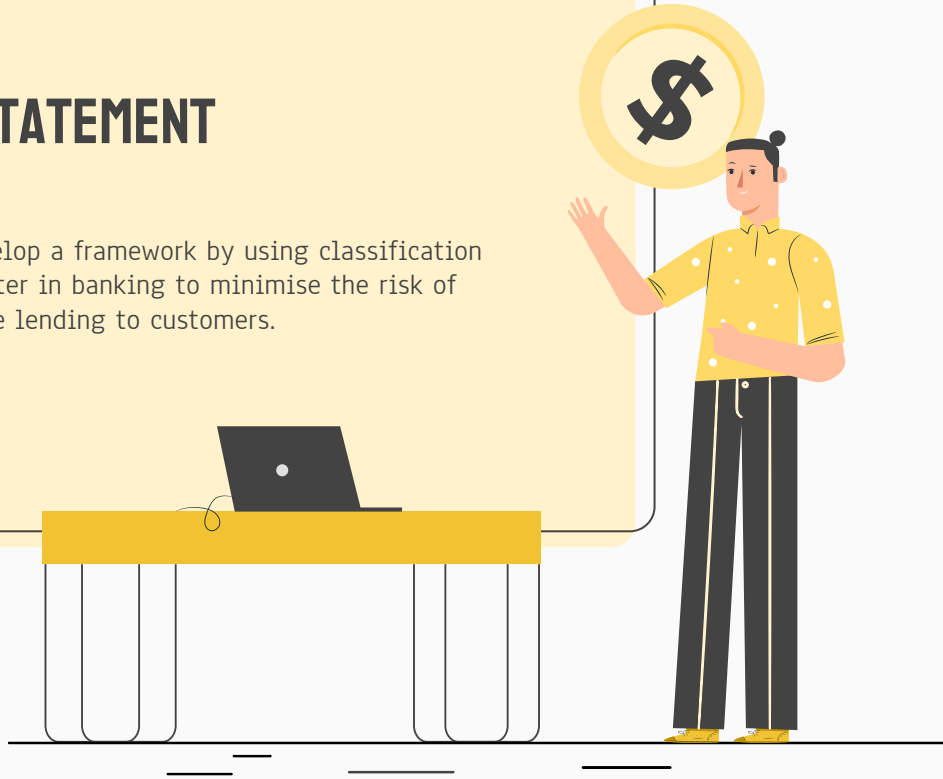
02
DESIGN & DATASET

03
RESULTS & ANALYSIS

04
CONCLUSIONS

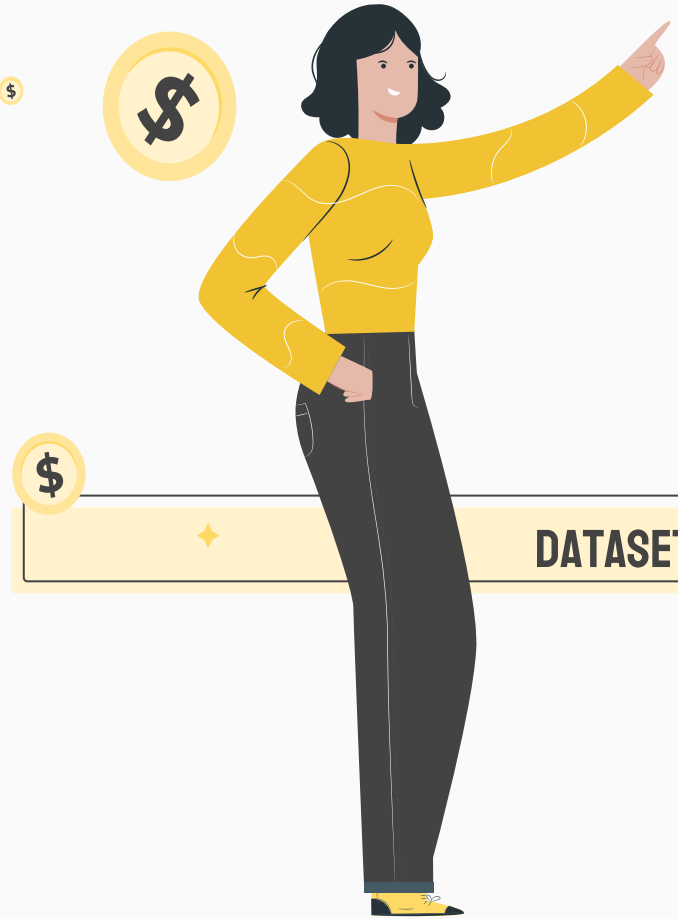
PROBLEM STATEMENT

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.



WHAT HAVE DONE

- 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

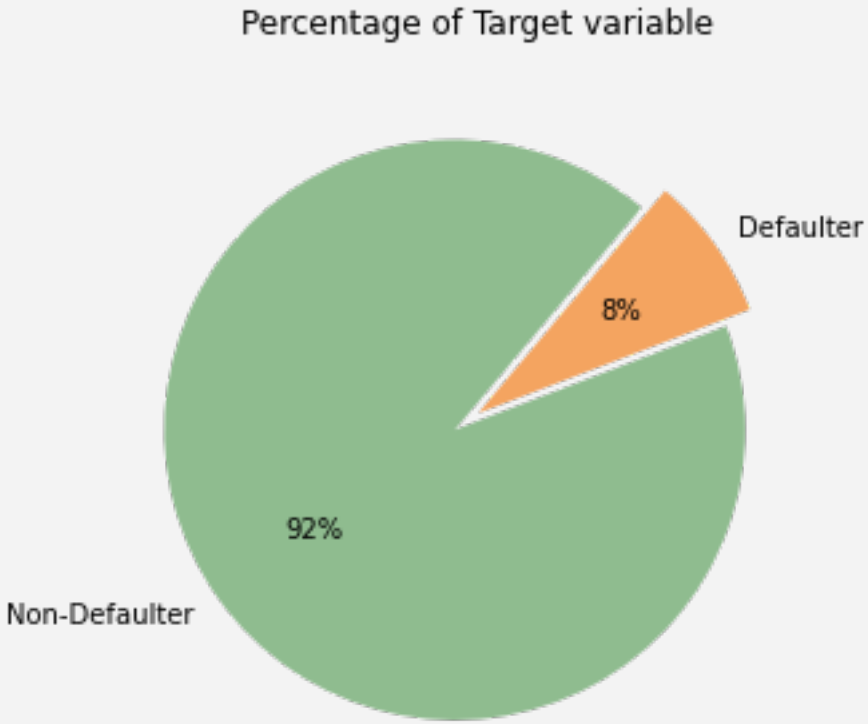
122

FEATURES

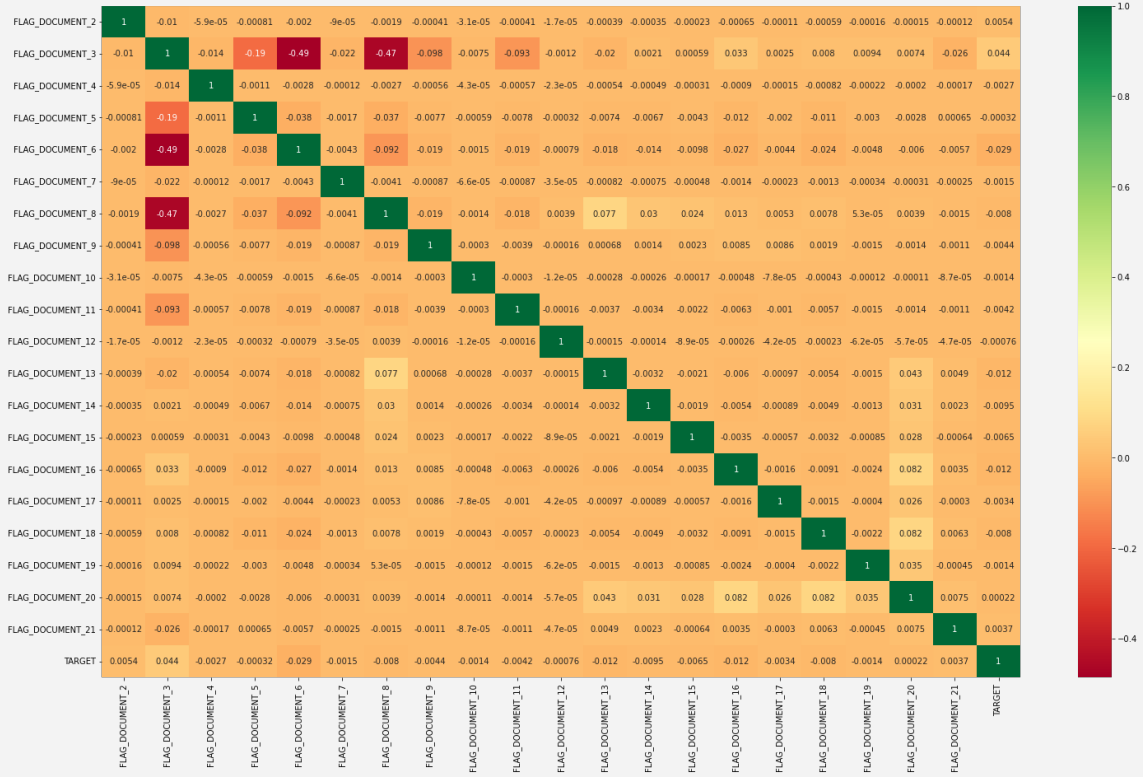
DATASET SUMMERY

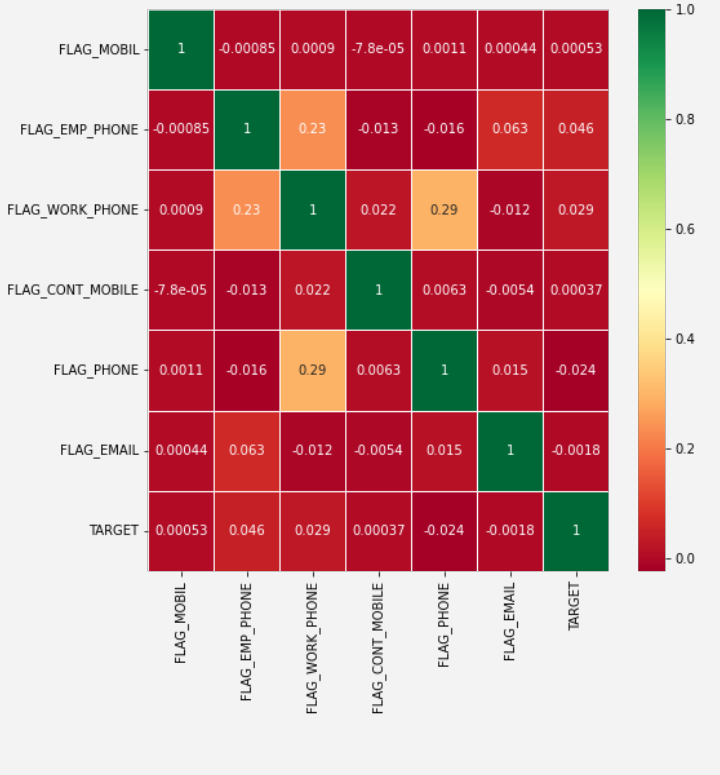
MODELS AND RESULTS

TARGET FEATURE DISTRIBUTION

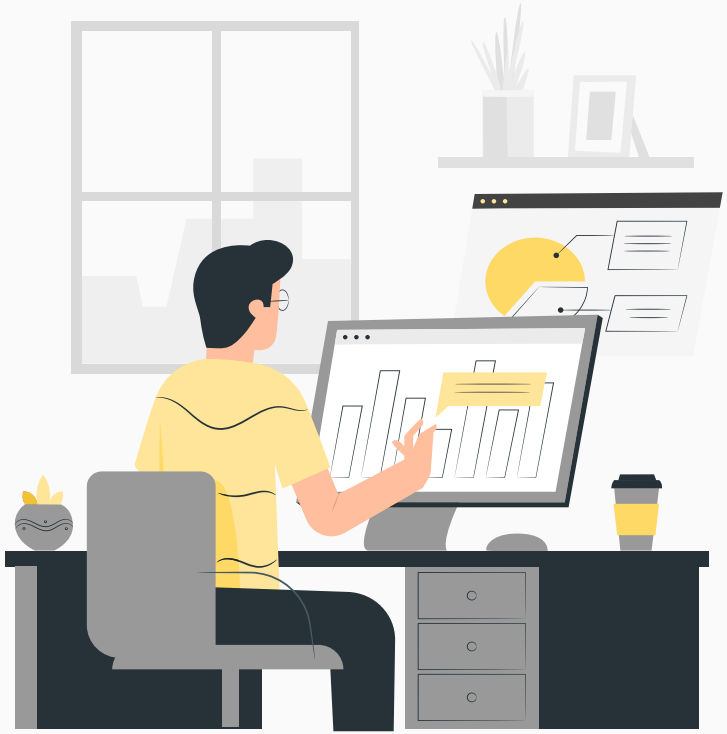


FEATURES CORRELATIONS





THE CLASSIFIERS MODELS



- ✓Logistic regression (LR)
- ✓K-nearest neighbours (K-NN)
- ✓Decision Tree (DT)
- ✓Random Forest (RF)

Confusion Matrix

Actual Values	Predicted Values: Positive	Predicted Values: 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

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?