VIRTUAL INTERNSHIP PROGRAM

DATA SCIENTIST

IMAN ABDURRAHMAN

HOME CREDIT DEFAULT RISK PREDICTION

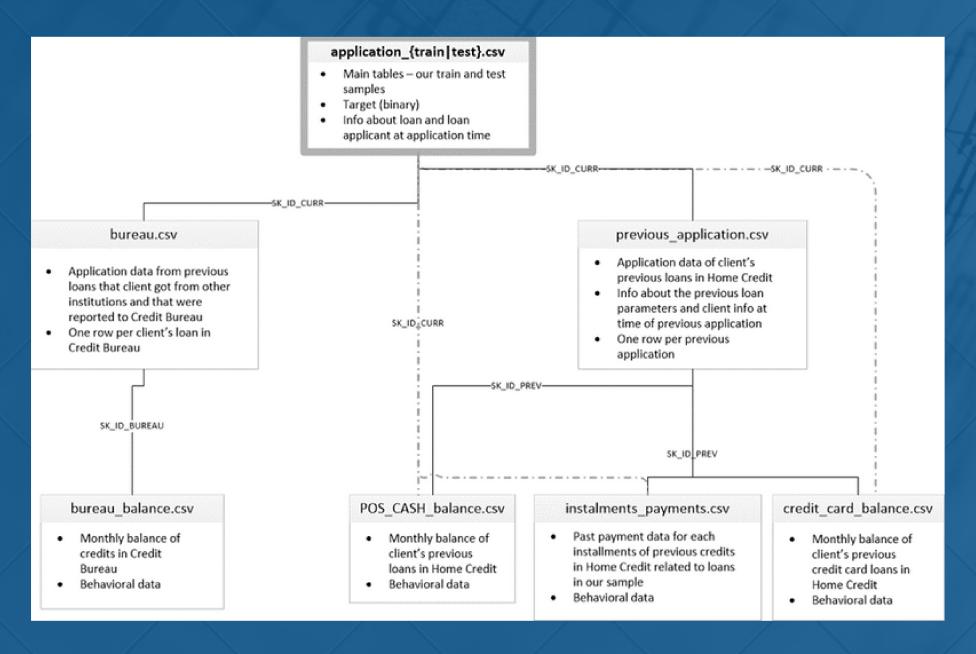


id/x partners

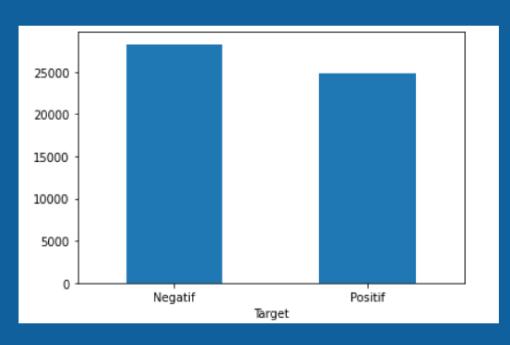
Problem Research

Many people struggle to get loans due to insufficient or nonexistent credit histories. Home Credit makes use of a variety of alternative data to predict their clien's repayment abilities

DATA DESCRIPTION



DATA DESCRIPTION

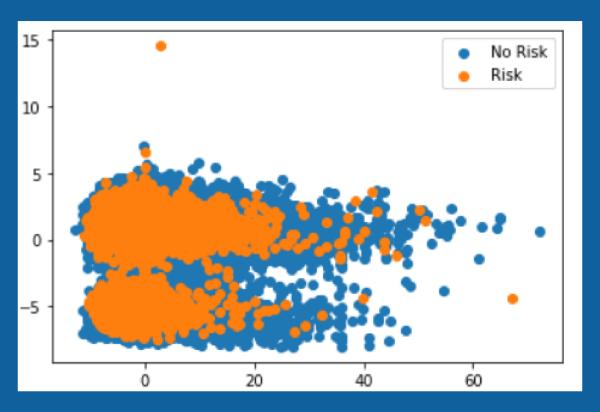


TOTAL NO. OF FEATURES: 121
TOTAL NO. OF OBSERVATIONS: 304531
ACCURACY PARADOX
IMBALANCED CLASS:

• MORE LOANS WERE REPAID THAN LOANS THAT WERE NOT REPAID

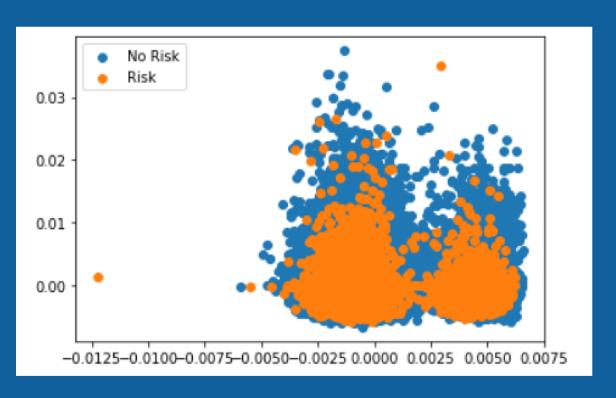
DATA VISUALIZATION

PCA

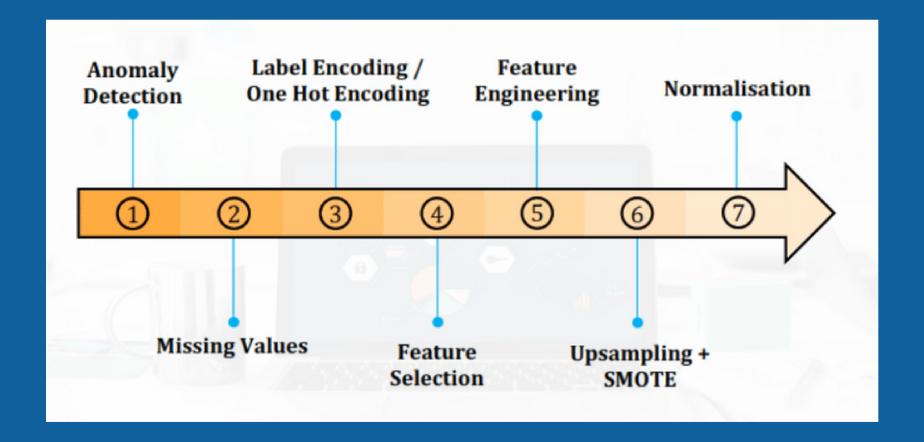


DATA VISUALIZATION

ICA



DATA PREPROCESSING



MODELLING

Algorithm	Accuracy	F1-Score	ROC AUC
Logistic Regression	91.92 %	0	0.62
Decision Tree	85.54 %	0.15	0.54
Naive Bayes	91.2 %	0.012	0.60

Conlusion

- Knowing the complexity of the dataset
 - Missing values, multiple sources
- Imbalance Dataset
 - Accuracy Paradox
 - Insensitivity to minority class, resulting in poor F1-Score