NOV 22, 2019

HOME CREDIT DEFAULT

Optimized Consumer Lending via Machine Learning

Presentation Outline

Team
Business Case
Data Exploration
Predictive Model
Business Recommendation
Future Work

Team

Frische Fische

GESCHE WEGENER

Economist

LENNART JUNGE

Investment Manager

Business Case

OPTIMIZED CONSUMER LENDING VIA MACHINE LEARNING

PREDICTIVE MODELLING

Classifying the consumers who will pay back their credits and the consumers defaulting

FEATURE IMPORTANCE

Which features are most important in telling which lender will pay back its credit?

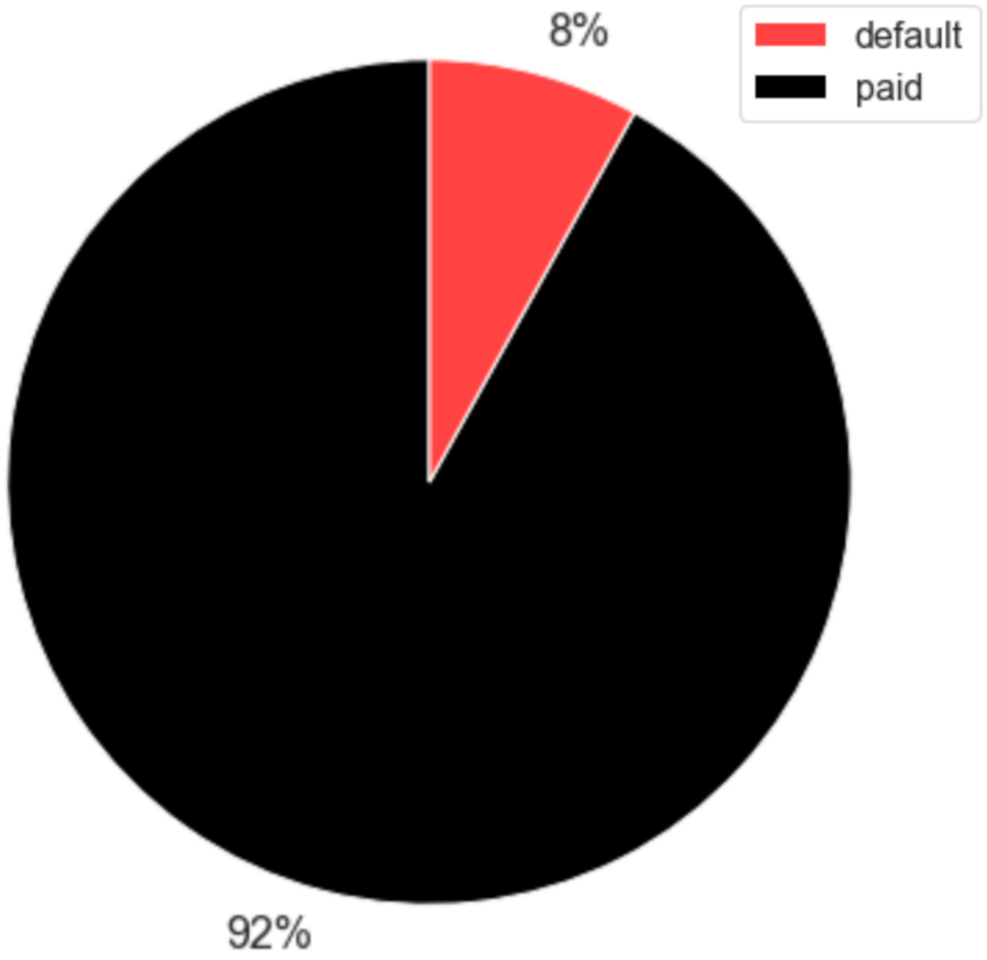
Data Exploration

DATA IMBALANCE

8% of borrowers default on the credit
92% of borrowers paid back the credit

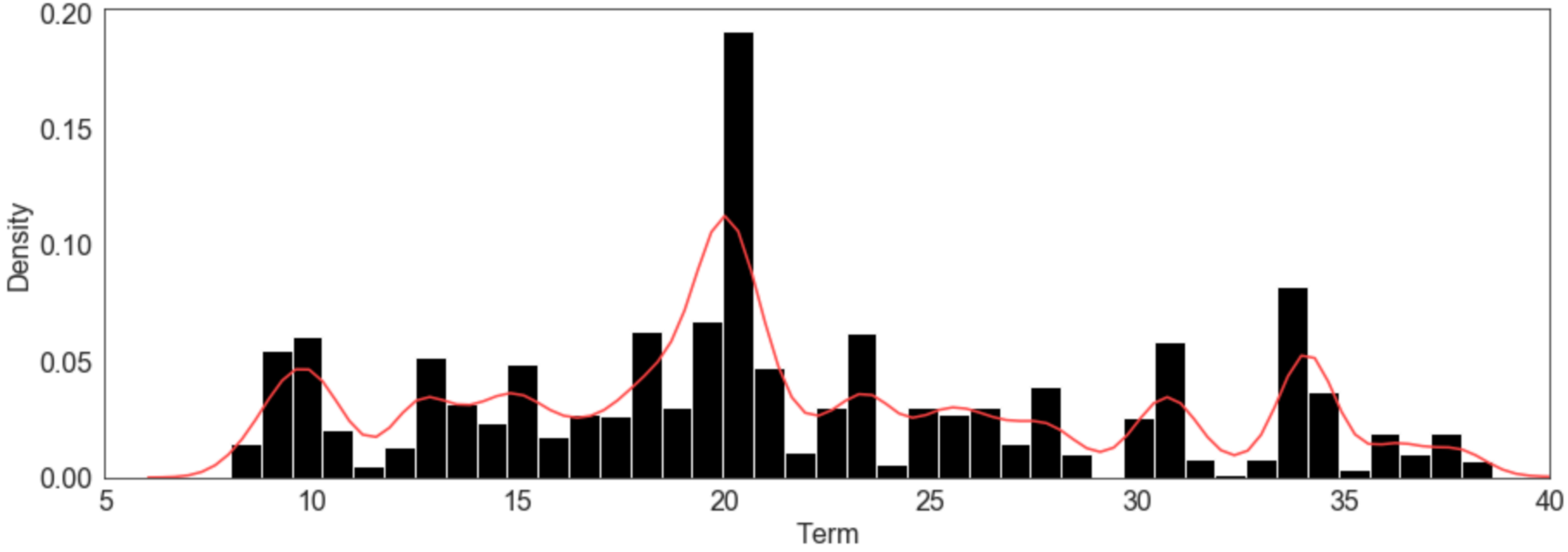
PERFORMANCE METRIC

Recall: How many of the defaulting borrowers are classified correctly?



Data Exploration

CREDIT TERM

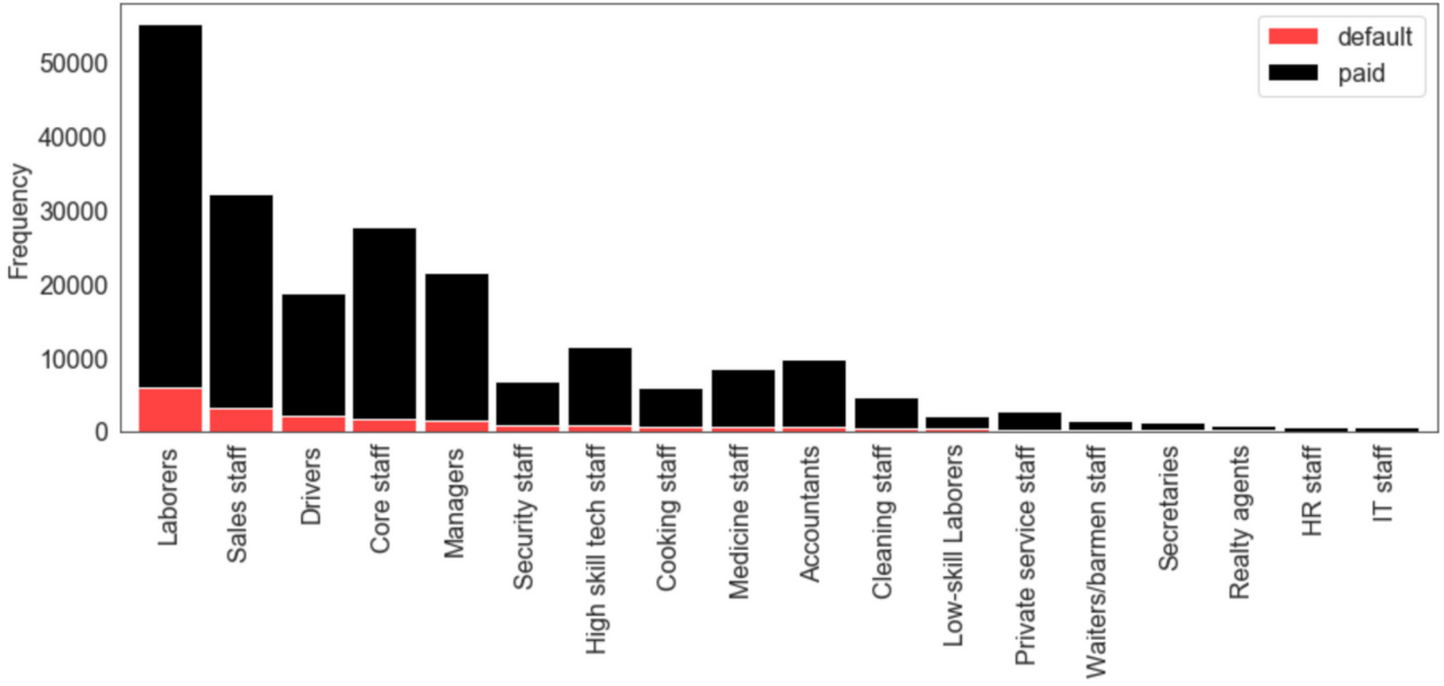


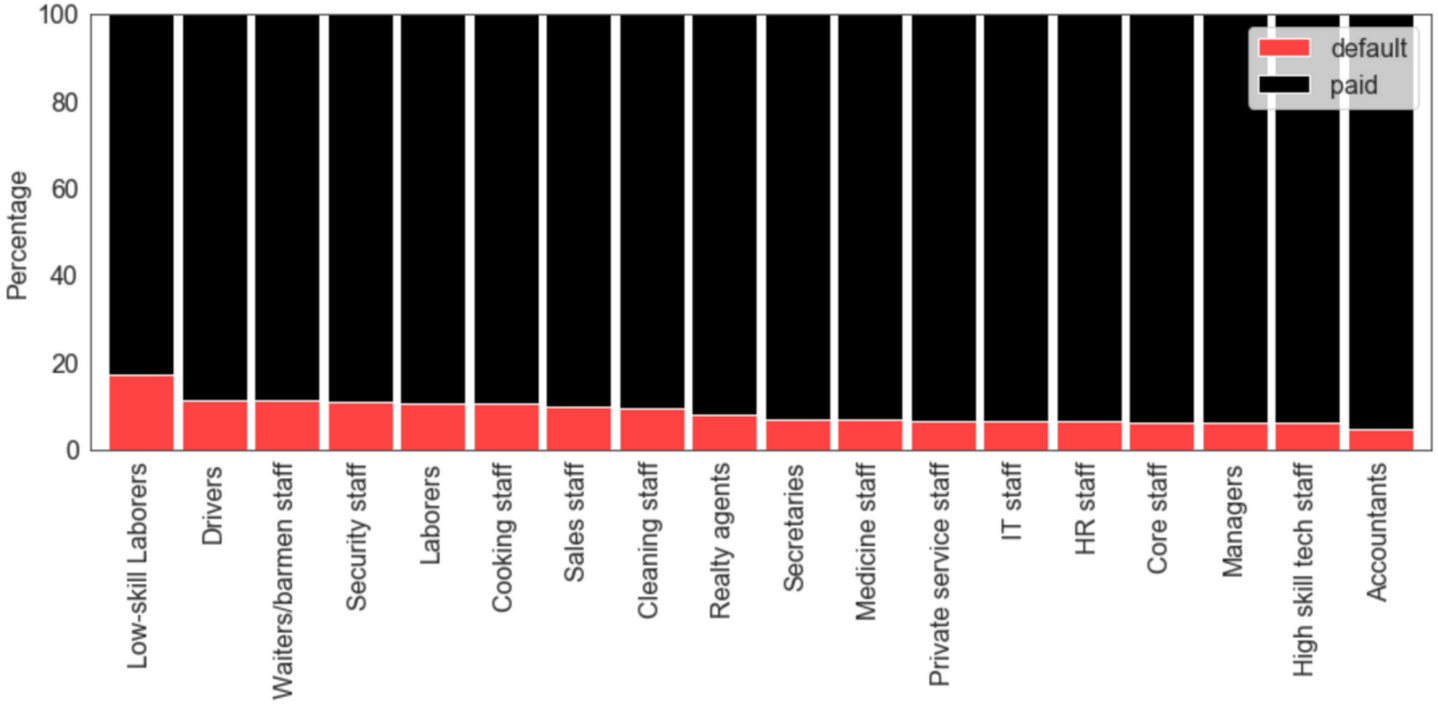
Data Exp.

OCCUPATION

Most of the borrowers are laborers and other low skilled staff

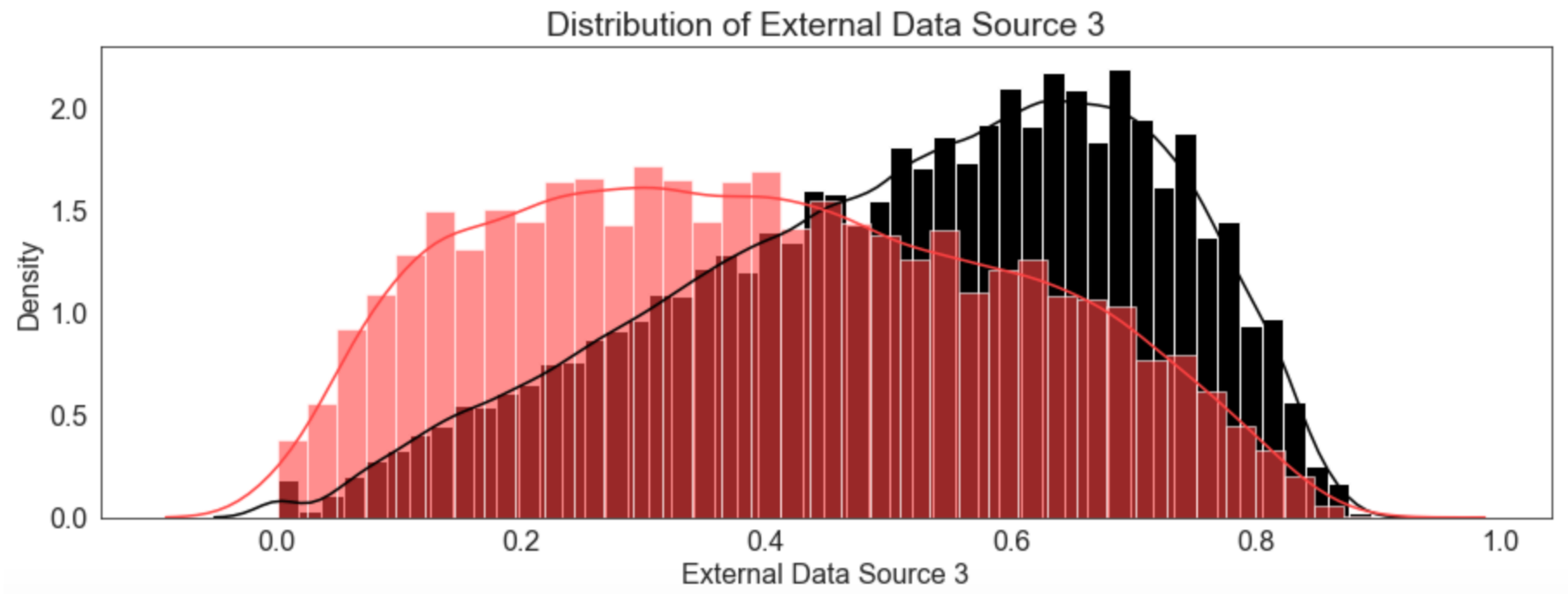
The percentage of borrowers defaulting on the credit is highest for low-skilled laborers





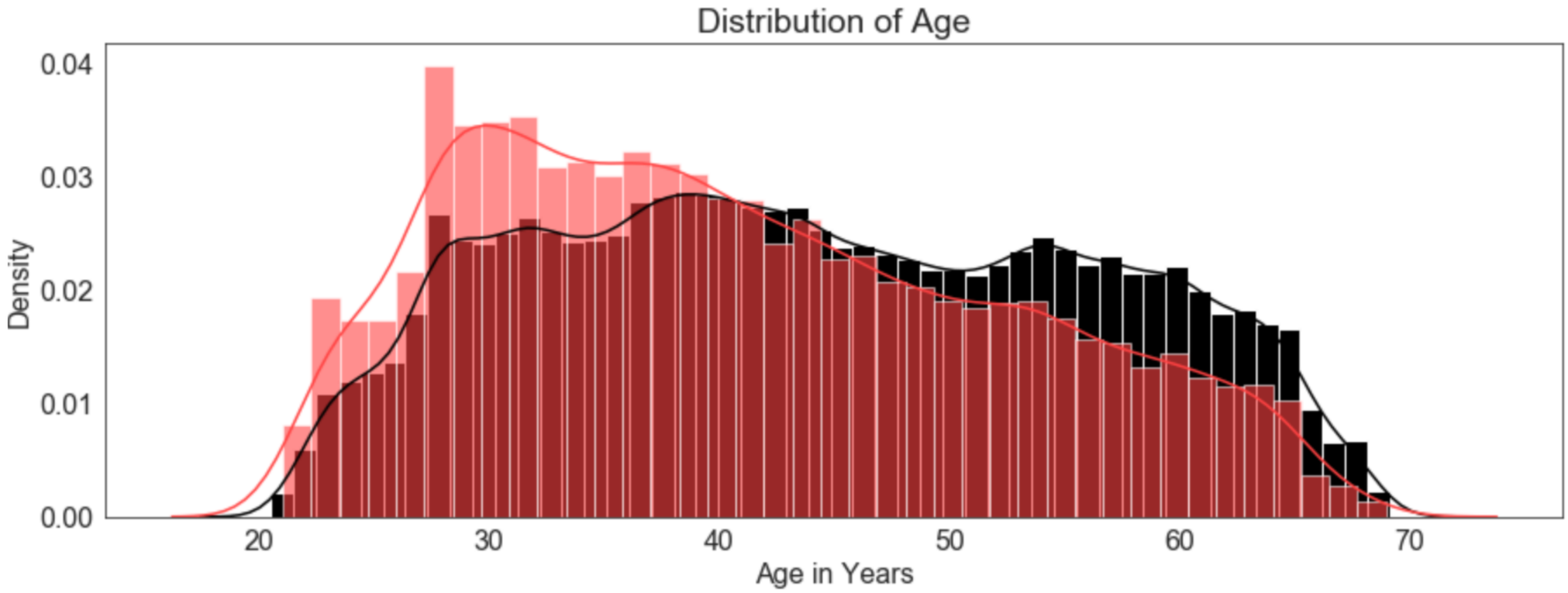
Data Exploration

CREDIT SCORE



Data Exploration

AGE



Data Exploration

YEARS OF EMPLOYMENT

Predictive Modelling

LOGISTIC REG

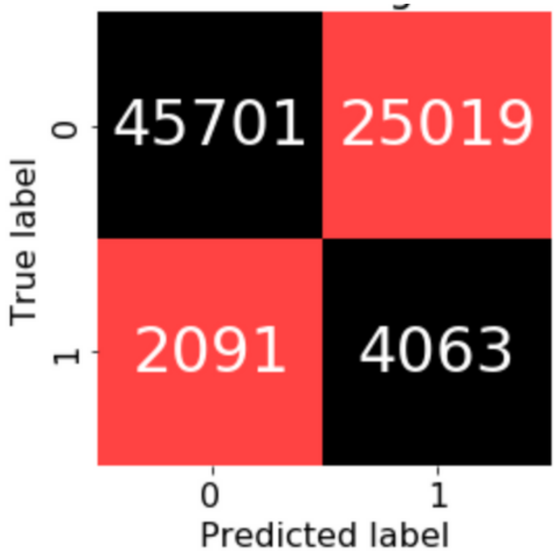
Recall: 65%

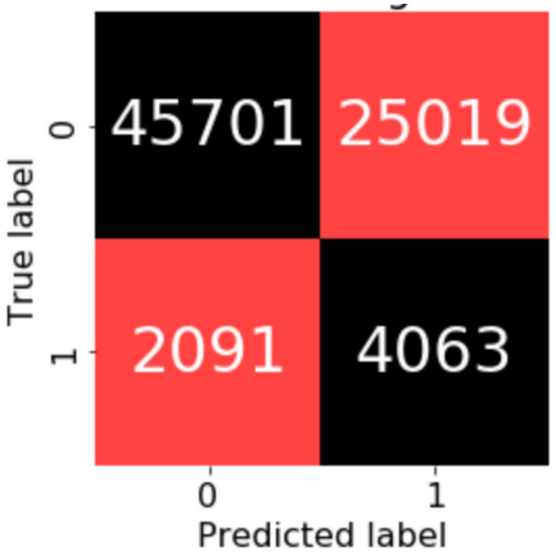
RANDOM FOREST

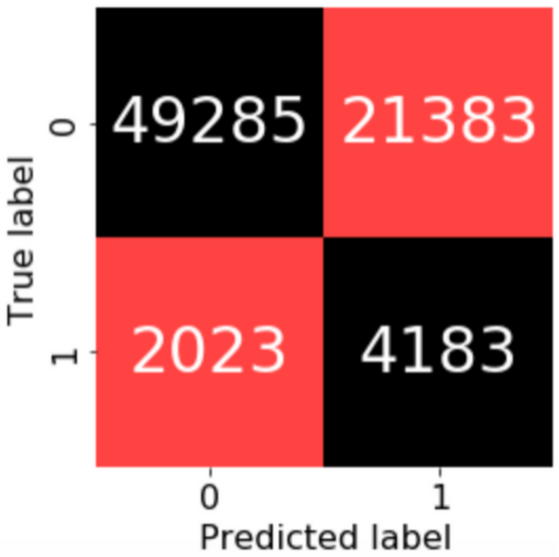
Recall: 88% (1: 27%)
Optimized Recall: 68%

ADABOOST

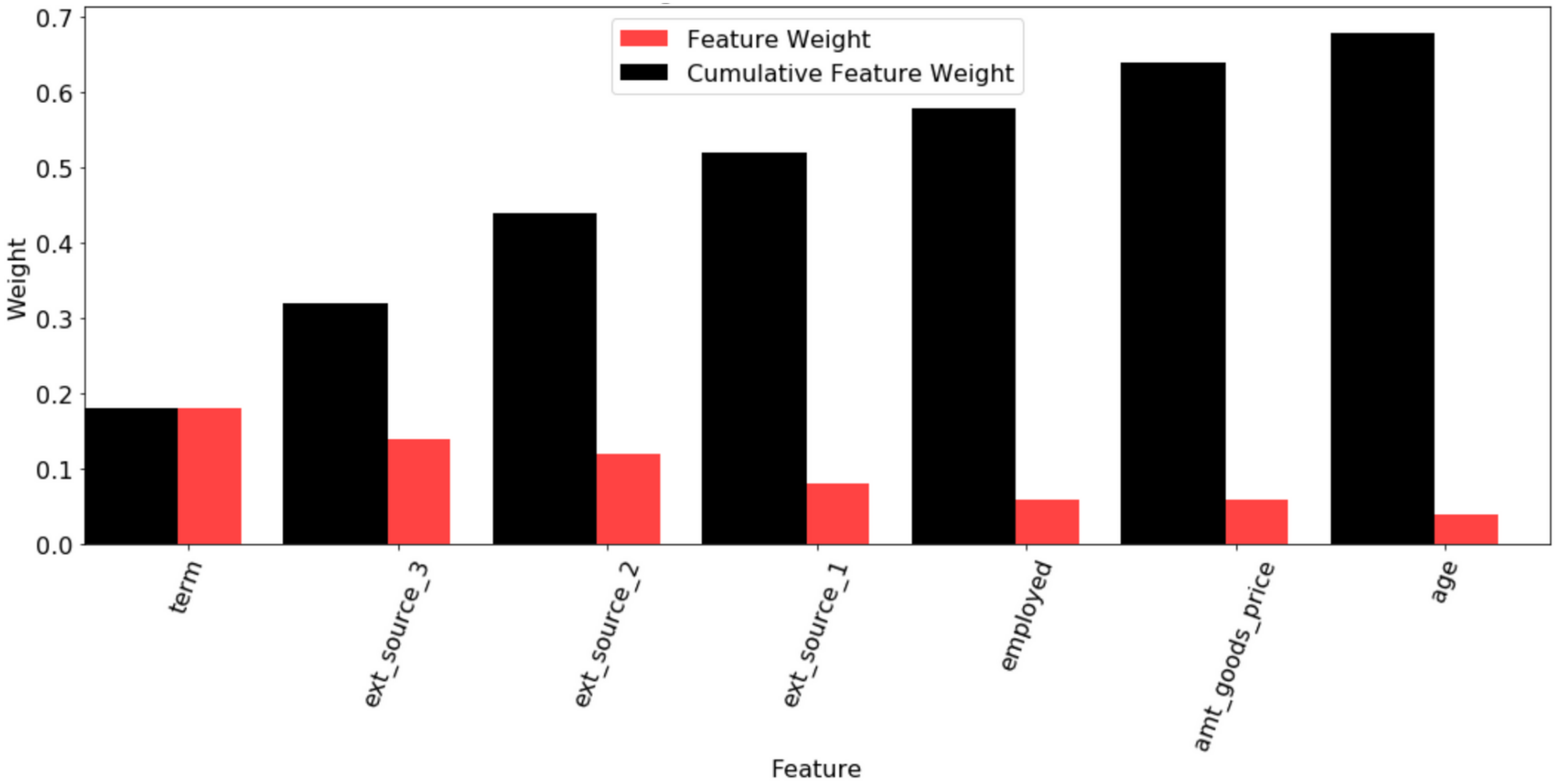
Recall: 70%
Optimized Recall: 72%







Feature Importance



Business Recommendation

CREDIT SCORES

external credit scores are a good indicator on the likelihood of a default

CREDIT TERM

longer credit terms tend to increase the likelihood of a default

AGE & EMPLOYMENT

younger borrowers and borrowers with lower years of employment tend to default more often

Future Work

ADDITIONAL DATA

Include additional Data Sets to reach a better model fit

FURTHER MODEL SPECIFICATION

Further train single models in order to reach higher recall score and ROC score

MODEL STACKING

Combining the optimized models of Logistic Regression Model, Random Forest, AdaBoostClassifier and potentially further models like XGBoost and SVM

THANK YOU

PRESENTED BY GESCHE & LENNART