Rakamin Academy

HOME CREDIT

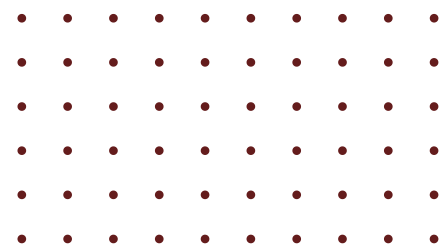# HOME CREDIT
CREDIT SCORING MODEL
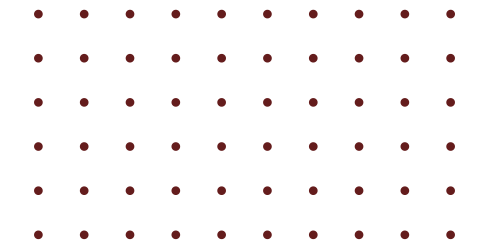
20 MARET 2023

agus abdul rahman

TABLE OF CONTENT

PROJECT BACKROUND

Home Credit currently uses various statistical methods and machine learning to predict credit scores. However, there is still a risk of being rejected by potential customers who are able to repay due to insufficient credit history or lack of alternative data. The problem is to identify potential customers who are at risk of default on loans and provide them with a positive loan experience.

PROBLEM RESEARCE

• Data Source

The data used are **application train** and **application test**. There are our main table, broken into two files for train (with TARGET) and test (without TARGET).

• Objective

1. **Develop** a machine learning model to identify potential customers who are at risk of loan default.
2. **Provide recommendations** for improving the loan application process for these customers, such as adjusting the principal, maturity date, and calendar payments to increase the likelihood of successful payments.

• Target

The goal is to increase financial inclusion and provide a positive lending experience for underserved populations. By identifying potential clients who are at risk of defaulting on loans and providing them with a positive loan experience, Home Credit can increase their clients' chances of success in applying for loans.
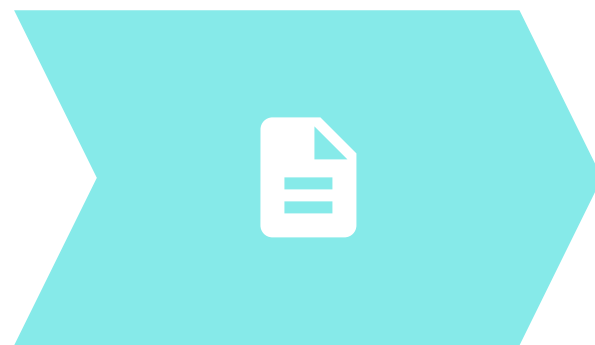
Business Metrics

1. Lower the loan default rate by 5% within the next year.
2. Increase the loan amount by 10% in the next year.
3. Increase the customer satisfaction rate by 5% in the next year.
4. Increase the proportion of customers who have successfully paid off their loans by 5% in the next year
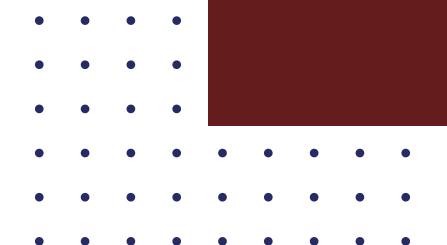
DATA PREPROCESSING

DATA APLICATION TRAIN

1 - EDA

- Bivariate Visualization
- Multivariate Visualization

2 - Data Cleaning

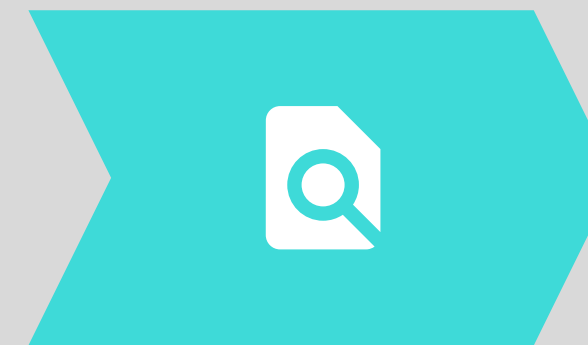- Detecting Duplication
- Handling Missing Values
- Detecting Outliers

3 - Model Building

- Label Encoding
- Feature Selection
- Handling Imbalanced Data
- Model Building
- Model Evaluation

DATA APLICATION TRAIN
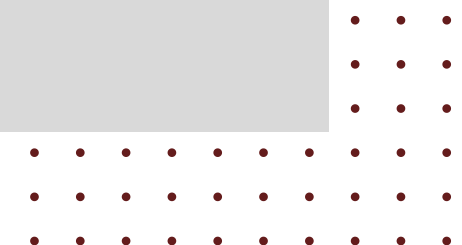
1 - Data Cleaning

- Detecting Duplication
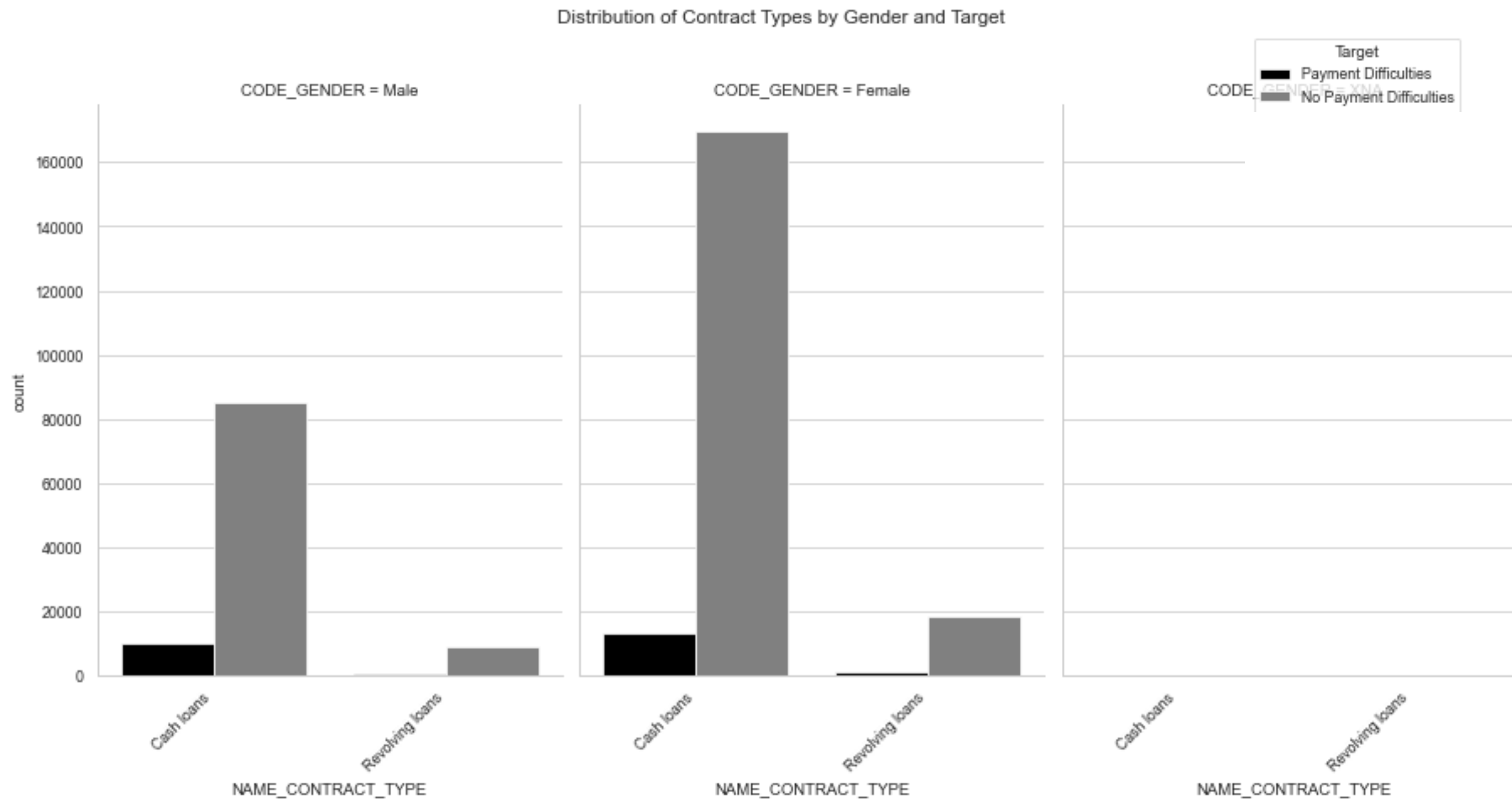- Handling Missing Values
- Label Encoding

2 - Prediction

- Predict client's repayment abilities with best machine learning model obtained before
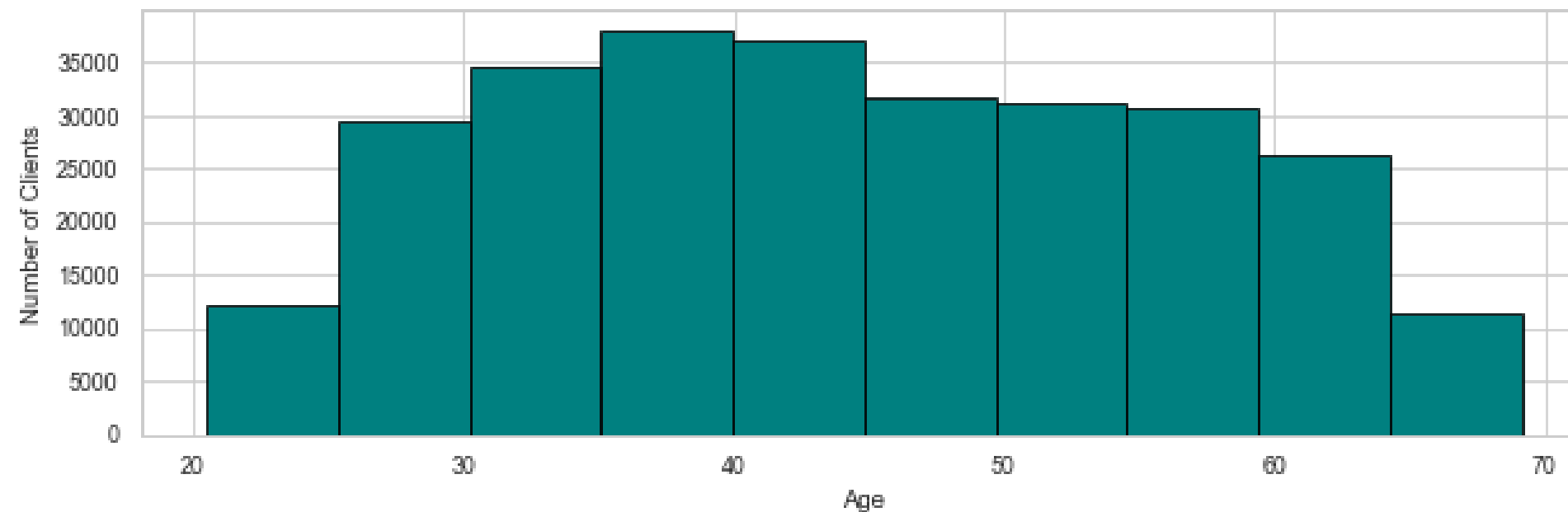
BUSINESS INSIGHT
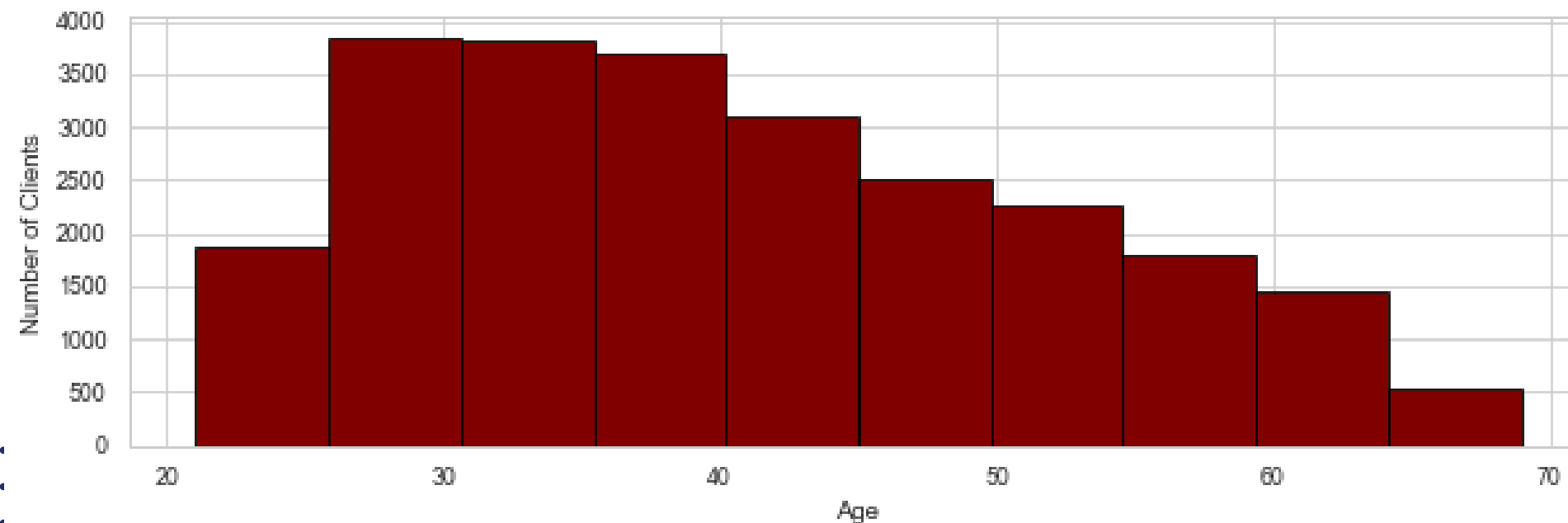

Distribution of Contract Types by Gender and Target

"From the given data, we can conclude that clients with **cash loans have a higher number of payment difficulties** as compared to those with revolving loans. It is also observed that **females have a higher number of payment difficulties** as compared to males. This information can be useful for financial institutions to make informed decisions while approving loans, and they can focus on reducing the risk associated with clients with cash loans and female clients."

BUSINESS INSIGHT

Age of Client (in years) who have No Payment Difficulties

AGE CLIENT INSIGHT

Clients aged **35-45** years are less likely to experience payment difficulties and could be prioritized as a target audience. Conversely, clients aged 25-35 years are more likely to face payment difficulties.

Age of Client (in years) who have Payment Difficulties

Target clients aged 35-45 years as they have fewer payment difficulties, whereas clients aged 25-35 years tend to experience more payment difficulties.
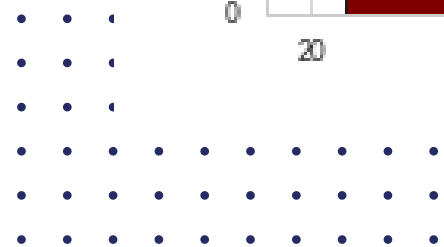
MACHINE LEARNING MODEL

MODEL COMPARISON

Algorithm	Training Accuracy Score	Testing Accuracy Score	Error Margin	ROC Score
Logistic Regression	67.49%	67.69%	32.31%	0.6769
Gaussian Naive Bayes	63.17%	63.52%	36.48%	0.6353
Decision Tree	100%	89.63%	10.28%	0.8963
Random Forest	**100%**	**99.73%**	**0.27%**	**0.9973**
Gradient Boosting	67.72%	68.6%	31.4%	0.686
K-Nearest Neighbor	91.71%	88.14%	11.86%	0.8813

In conclusion, the Random Forest algorithm is the best choice for model classification based on the given data because it provides more accurate and reliable results in predicting new data.

BEST MODEL ALGORITHM RANDOM FOREST CLASSIFIER
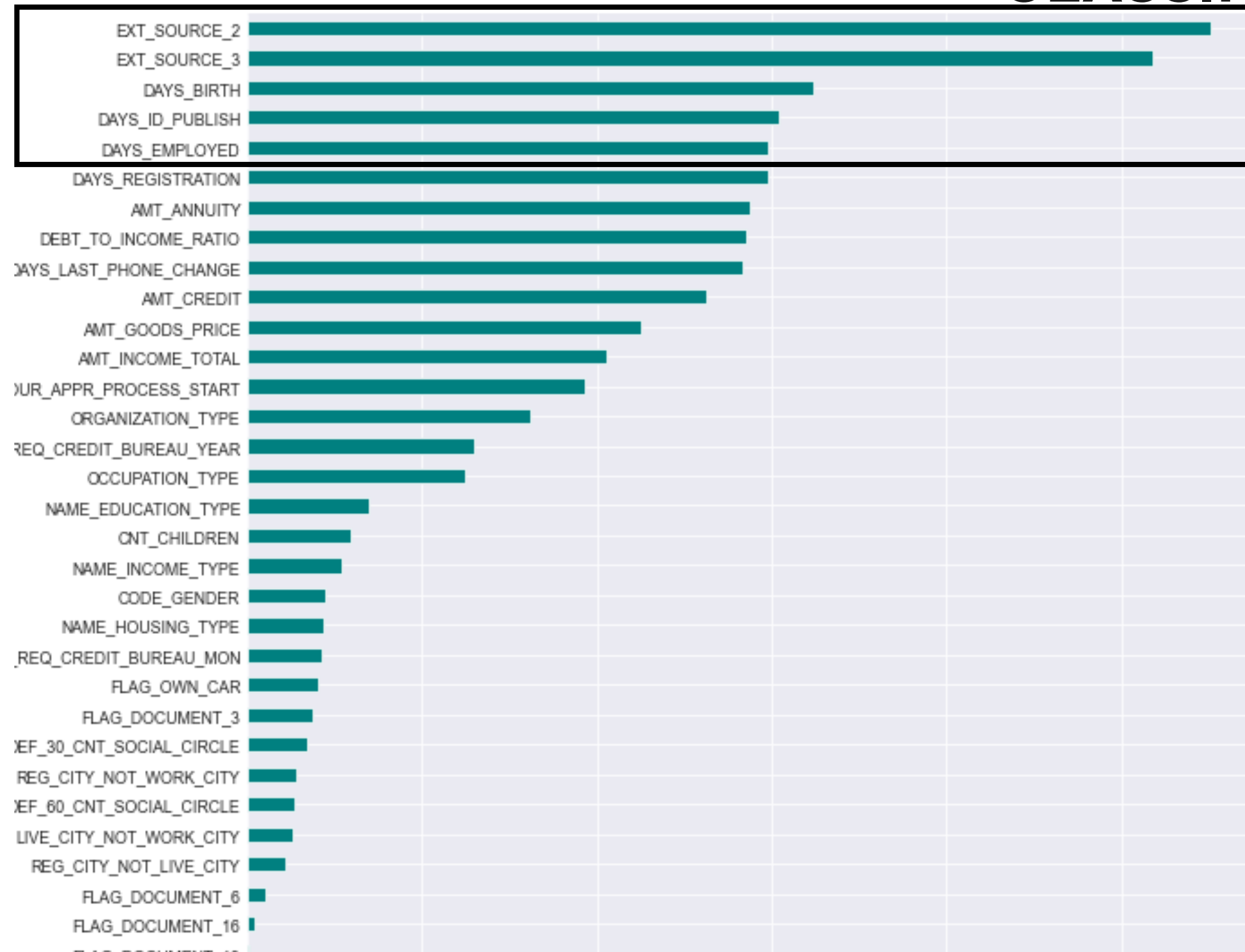
Performance:
- Random forest model gives 100% correct results
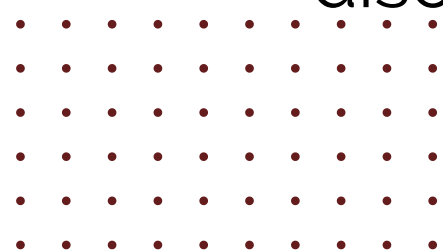- There is 0.27 % error margin

BUSINESS RECOMMENDATION

RECOMENDATION 1

Prioritize clients aged 35-45 years old, as they are less likely to experience payment difficulties, and develop targeted marketing strategies to attract these clients.

ACTION

1. Adjust marketing efforts to focus on attracting clients aged 35-45 years old through channels that are most likely to reach them, such as social media, targeted email campaigns, or sponsored events.
2. Offer personalized loan packages with terms and conditions that appeal to this age group, such as longer payment periods or lower interest rates for larger loans.
3. Provide incentives for existing clients in this age group to refer new clients, such as discounts or rewards for successful referrals.

RECOMENDATION 2

Implement a partial automation loan application process that can immediately accept or reject submissions based on the output of the model, and for submissions that are 'gray,' have them checked manually by the assessment team to ensure accuracy. This will increase the efficiency of the loan application process and reduce the risk of default.

ACTION

1. Develop and implement the partial automation loan application process with built-in models that assess a variety of factors that contribute to credit risk, such as credit history, income, and debt-to-income ratio.
2. Train the assessment team to accurately review and verify applications that are marked as 'gray' by the automated system, and ensure that they follow a consistent and fair review process.
3. Monitor the performance of the automated system regularly and refine the models to improve accuracy and reduce errors over time.
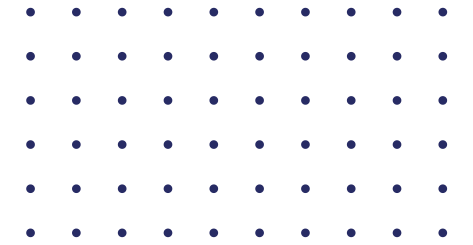
THANK YOU

https://github.com/agusabdulrahman/
Credit-Score-Card-Model