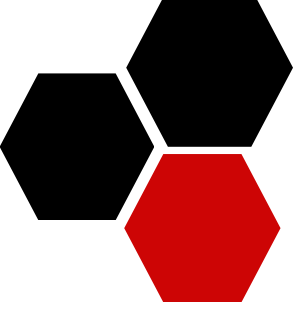# HOME CREDIT
SCORECARD MODEL

Overview:
1. Problem Research
2. Data Pre-Processing
3. Business Insight
4. Machine Learning Models
5. Business Recommendation

The entire project documentation:
https://github.com/riskaamylatulaskiyah/Home-Credit-Indonesia-ScoreCard

R i s k a   A m y l a t u l   A s k i y a h

01 PROBLEM SOLVING

PROJECT BACKGROUND

PT Home Credit Indonesia is a finance company that provides financing services for customers who shop online or offline. This financing facility is very instrumental in solving financial problems for loyal customers. Home Credit seeks to expand financial inclusion for untouched communities to provide a safe and positive financing experience. to ensure they get a good and safe service. Home Credit optimizes their data to be able to **predict the payment ability of their clients**. By doing this, they can ensure customers who can make payments are not rejected when applying for a loan, and loans can be given a principal, maturity, and payment calendar that will motivate customers to succeed.

DATA SOURCE

The data used are application_train.csv and application_test.csv. There is our main tablebroken into two files for the train (with TARGET) and test (without TARGET).

OBJECTIVES

1. Identify characteristics of potential clients who will have difficulty repaying loans.
2. Predict the client's repayment abilities.

CHALLENGES

1. Huge amount of data
2. Data is in imbalance
3. There is a missing value
4. There are outliers.

METHODOLOGY

1. Data Preparation and EDA for business insights.
2. Data Pre-Processing and Data Cleaning
3. Predictive Model Building using machine learning model.
4. Provide recommendations for the company to increase their client's success in applying for loans.

02 DATA PRE-PROCESSING

Data Application Test

121 columns

48,744 rows

DATA CLEANING

Detecting Duplication
0 data
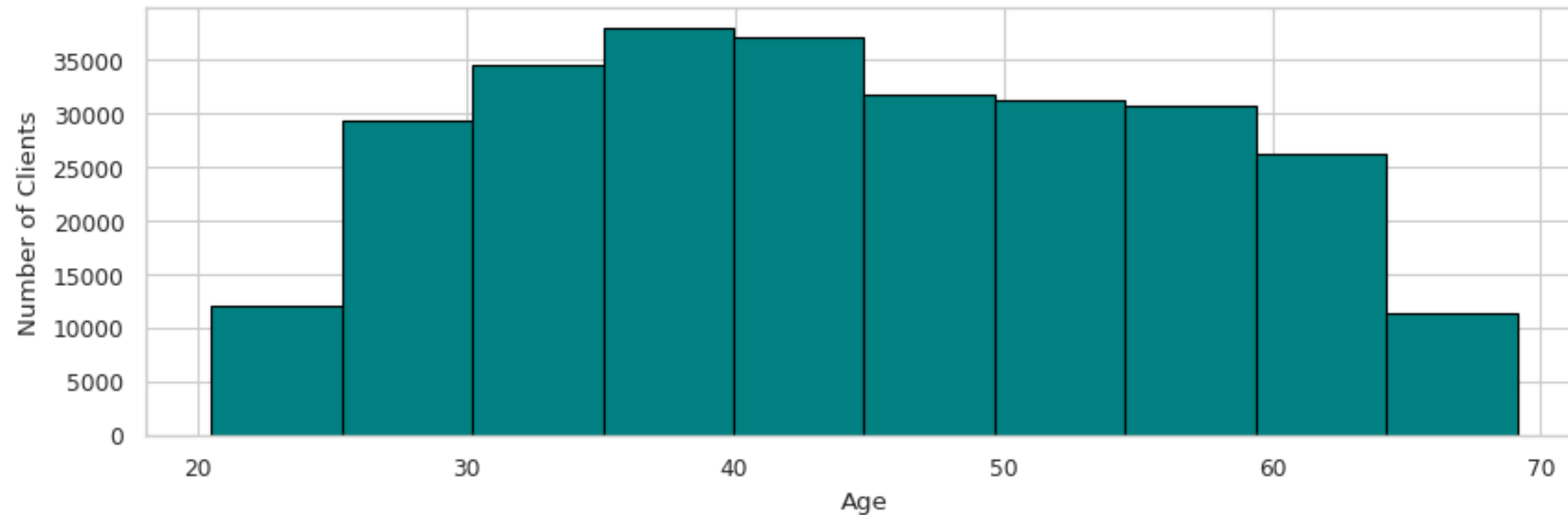
Handling Missing Values
67 columns

Label Encoding
Transform 12 non-numerical columns to numerical labels

PREDICTION

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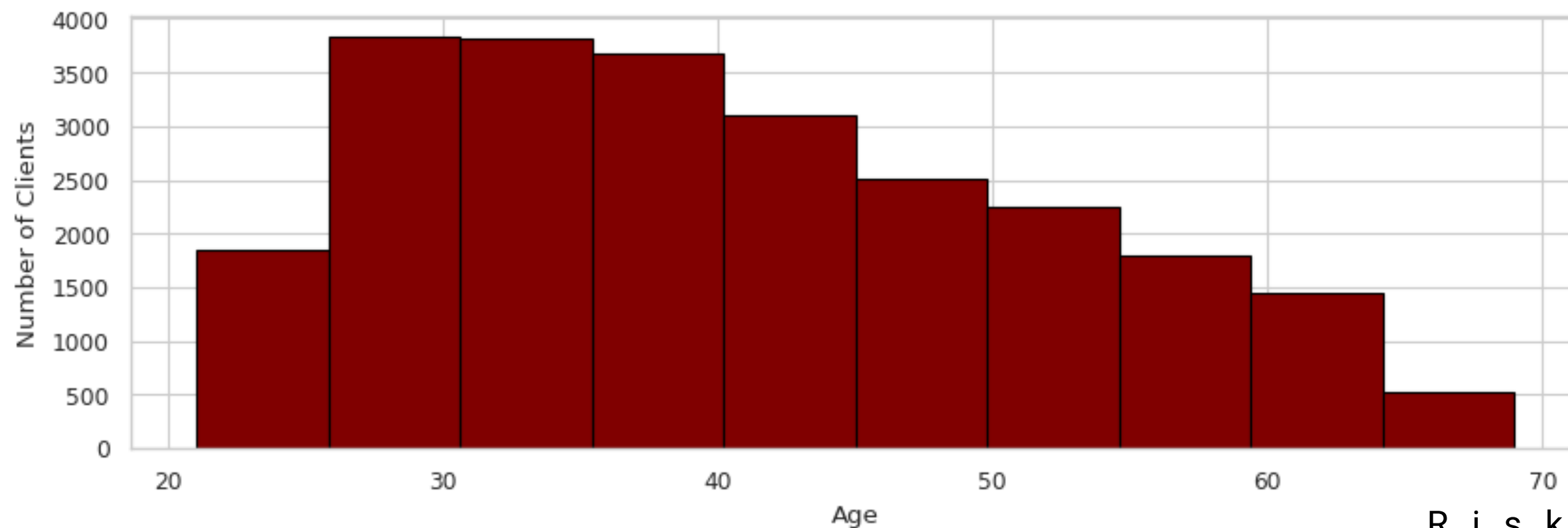
03 BUSINESS INSIGHT

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

- The distribution of data on the ages of customers who apply for loans is more dominant in the age range of 25-60 years.
- Within this age range, the age group that submitted the most requests was customers aged 35-40 years.

Age of Client (in years) who have Payment Difficulties

- Clients with no payment trouble are 35-45 years old. You can target this client as your own priority.
- Temporary clients with payment difficulty are in the 25-35 year range.

R i s k a   A m y l a t u l   A s k i y a h
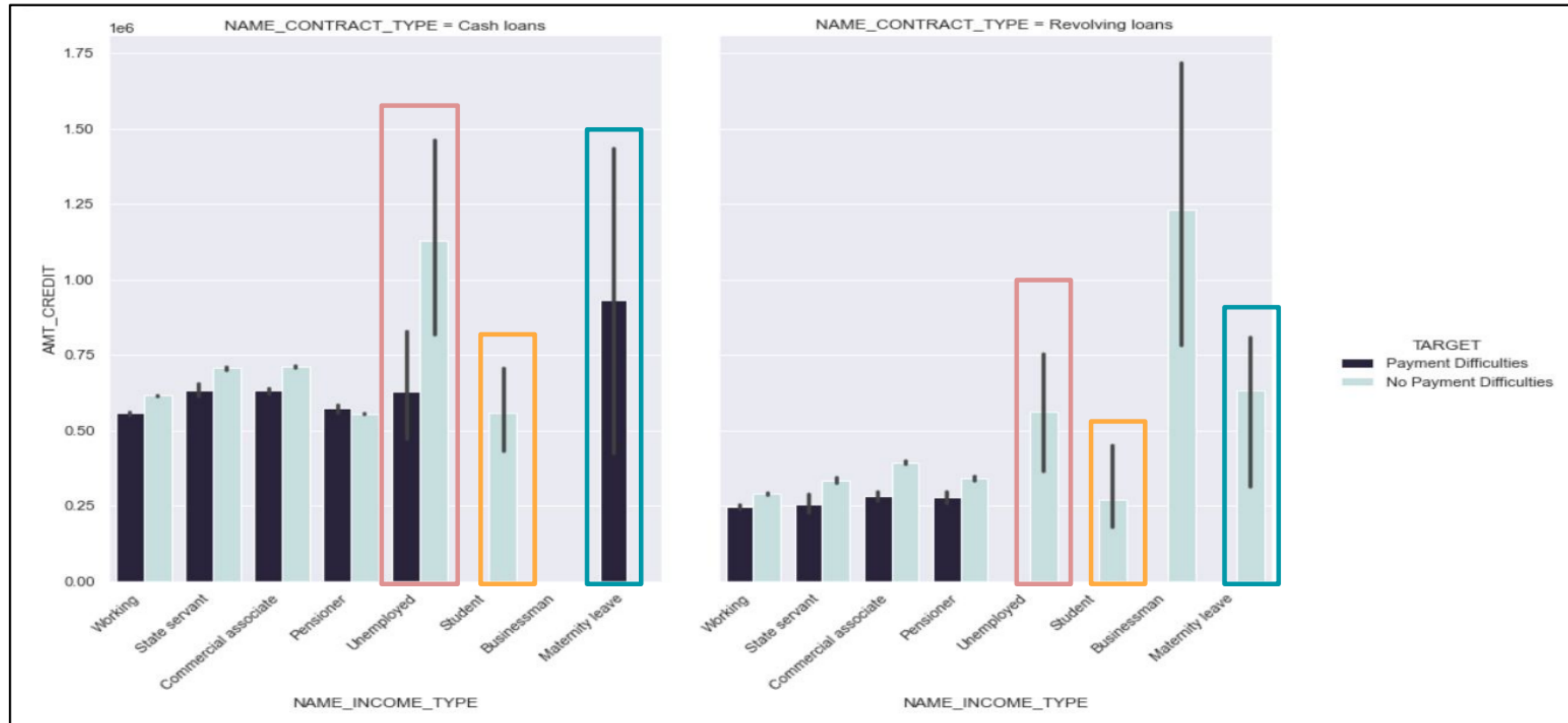
For unemployed clients with cash loans, more than 50% of clients have difficulty repaying loans with medium-credit loan amounts. While all unemployed clients with revolving loans have no difficulty returning loans.

For this type of maternity leave income with cash loans, all clients have difficulty repaying loans with medium loan amounts. Meanwhile, all clients with maternity leave and revolving loans have no difficulty paying off their loans.

All student clients have no difficulty repaying the loan. either with a cash loan or revolving loans for low to medium-credit loan amounts.

04 MACHINE LEARNING MODELS

Algorithm	Training Accuracy Score	Testing Accuracy Score	Error Margin	ROC Score
Logistic Regression	67.16%	67.29%	0.13%	0.6728
Gaussian Naive Bayes	60.24%	60.39%	0.15%	0.604
Decision Tree	100%	83.9%	11.74%	0.8826
Random Forest	100%	99.65%	0.35%	0.9965
K-Nearest Neighbor	91.56%	88.07%	3.79%	0.8806
Neural Network	70.01%	69.48%	0.58%	0.6948

The prediction accuracy of training and test data in the Random Forest model has a value that is not much different. This indicates that there is no underfitting or overfitting, so it concludes that the model is good. Therefore, the Random Forest model was chosen as the best model to predict the ability to pay clients.

Riska Amylatul Askiyah

Features Importance Plot
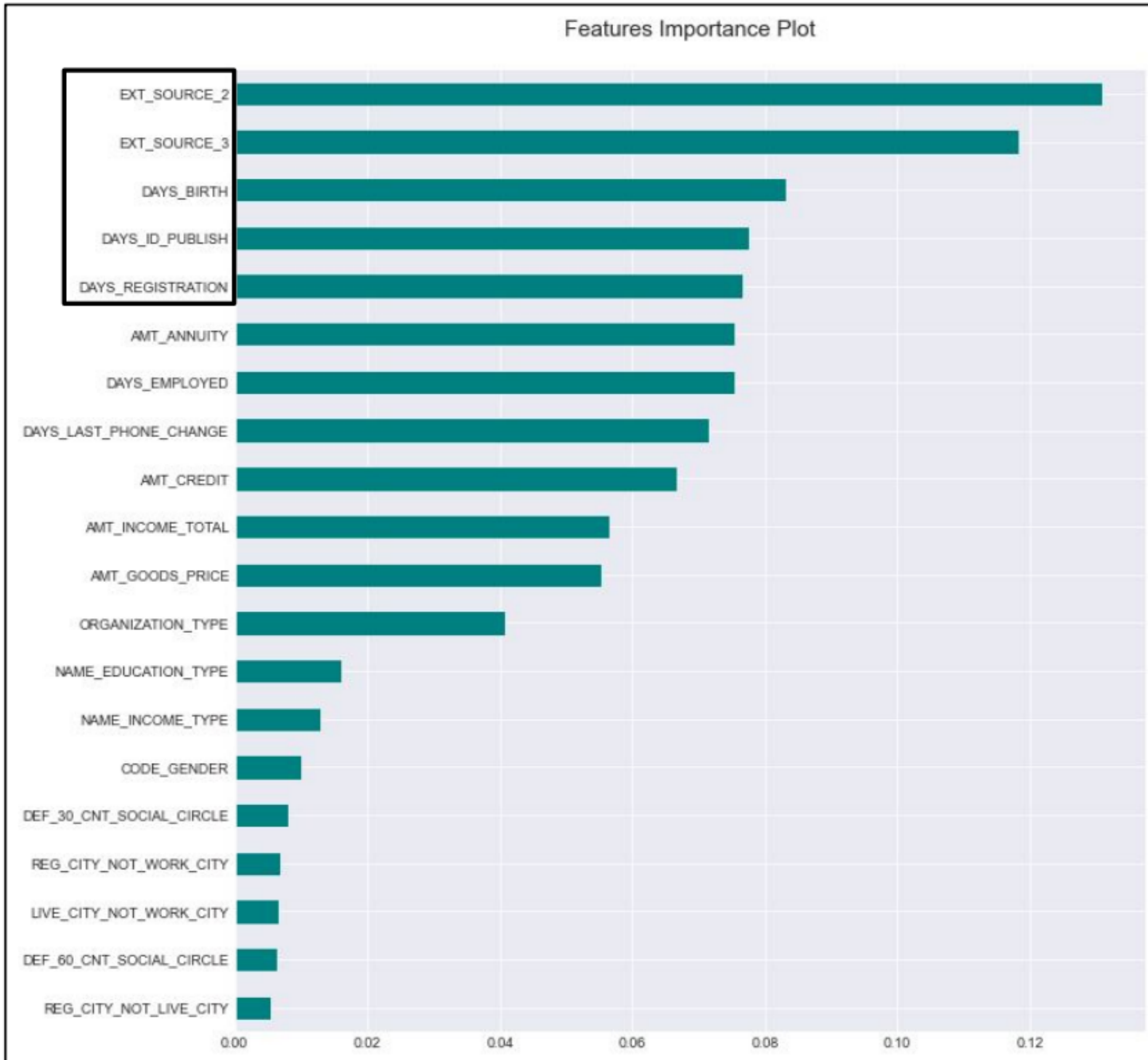
Algorithm	**Random Forest Classifier**

Performance	Random forest model gives **100% correct results**
	There is **0.35% error margin**

The 5 most important features
- Score from external data source 2
- Score from external data source 3
- Client's age in days
- Days ID publish
- Days registration

R i s k a   A m y l a t u l   A s k i y a h

RECOMMENDATION

- Clients with this type of student income are concluded to include clients who can pay off their loans either with cash loans or revolving loans (100% of applications approved). However, only 0.005% of applications come from students.
- Clients who work as accountants can be said to be able to return client loans (95% of applications are approved). However, only 3.19% of applications came from accountants. Similarly, clients who work as highly skilled technical staff and managers, can repay loans, but few applications come from them.

Create a campaign so that **more** student, accountant, high skill tech staff, manager **interested in applying for a loan**

Riska Amylatul Askiyah

RECOMMENDATION

- Clients with maternity leave and cash loans concluded include clients with payment difficulties (100% of applications rejected). Instead, all of the clients were on maternity leave but took out revolving loans to get their applications approved.
- More than 50% of unemployed clients have problems repaying their loans if they take cash loan contracts. Meanwhile, all unemployed customers who participated in the loan rotation were able to pay off their loans.

Need further analysis, you can **survey** to find out if there is a problem if a client with maternity leaves or unemployed takes a cash loans contract. So, in the future, if there are clients with that type of income, you **can recommend the right contract type** so that their applications will be approved

The entire project documentation: https://github.com/riskaamylatulaskiyah/Home-Credit-Indonesia-ScoreCard

R i s k a   A m y l a t u l   A s k i y a h