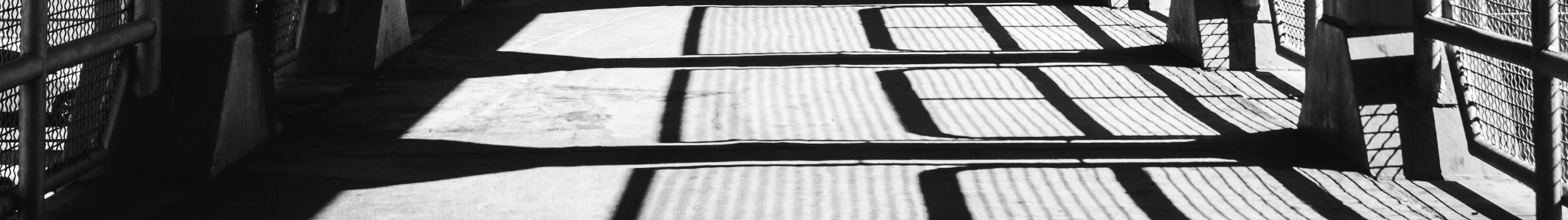




HOME CREDIT SCORE CARD MODEL

Made By Yudha Mulyana





Daftar Isi

- ```
graph LR; 1[1 Problem Research] --> 2[2 Data Preprocessing]; 2 --> 3[3 Business Insights]; 3 -.-> 4[4 Machine Learning Model]; 4 --> 5[5 Business Recommendation]; 5 --> 6[6 Feature Importance];
```

1 Problem Research

2 Data Preprocessing

3 Business Insights

4 Machine Learning Model

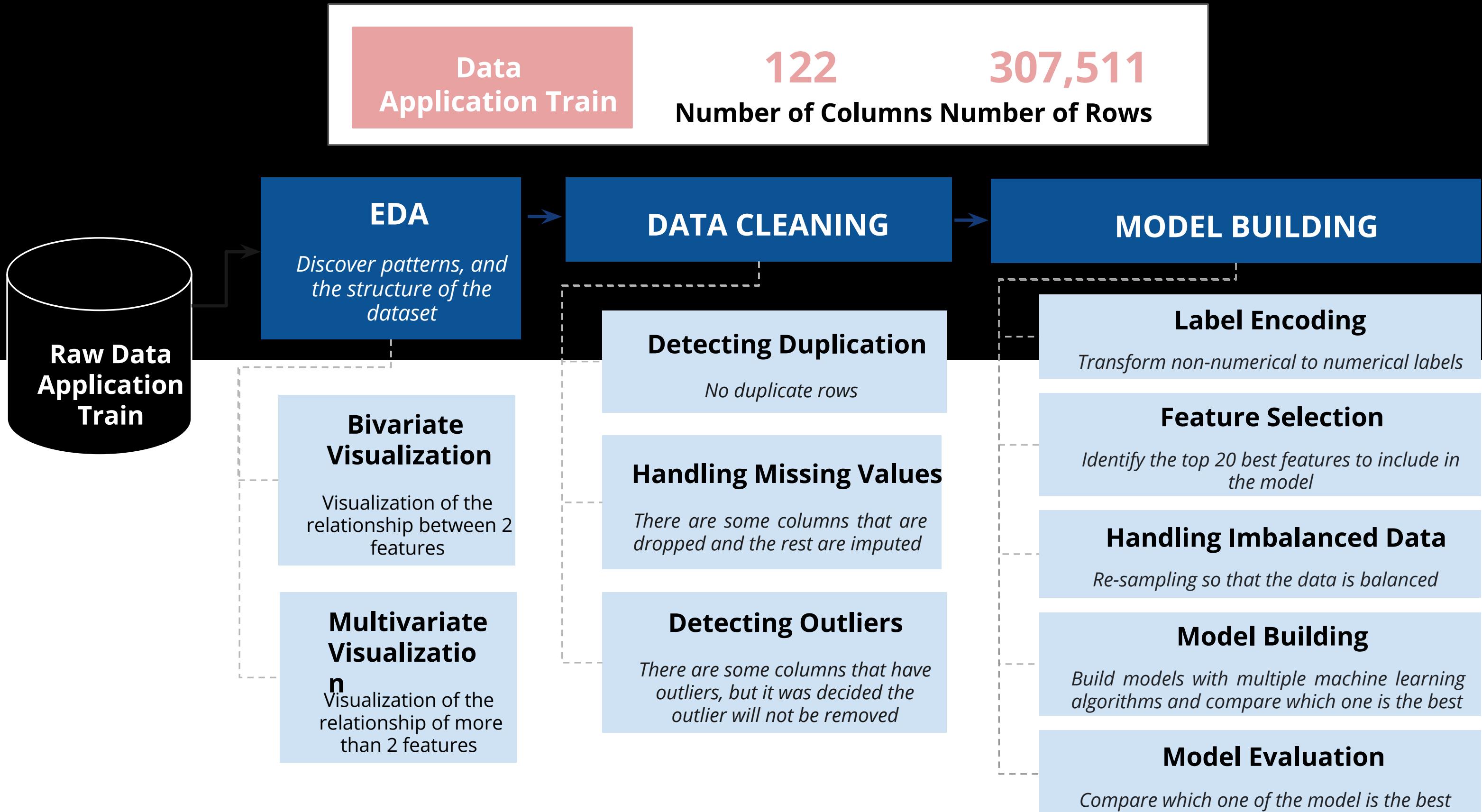
5 Business Recommendation

6 Feature Importance

# Project Background

Many people struggle to get loans due to insufficient or non-existent credit histories. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience. Home Credit makes use of a variety of alternative data to predict their clients' repayment abilities. Doing so will ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful





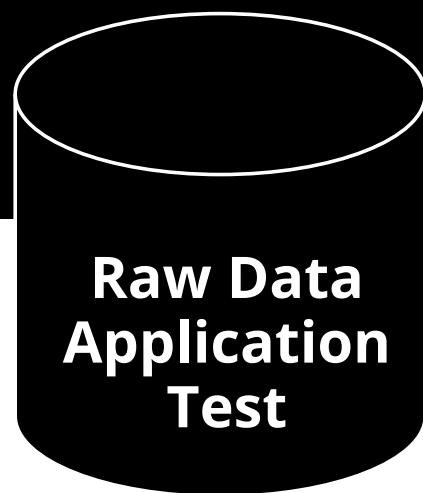
## Data Application Test

121

Number of Columns

48,744

Number of Rows



## DATA CLEANING

### Detecting Duplication

No duplicate rows

### Handling Missing Values

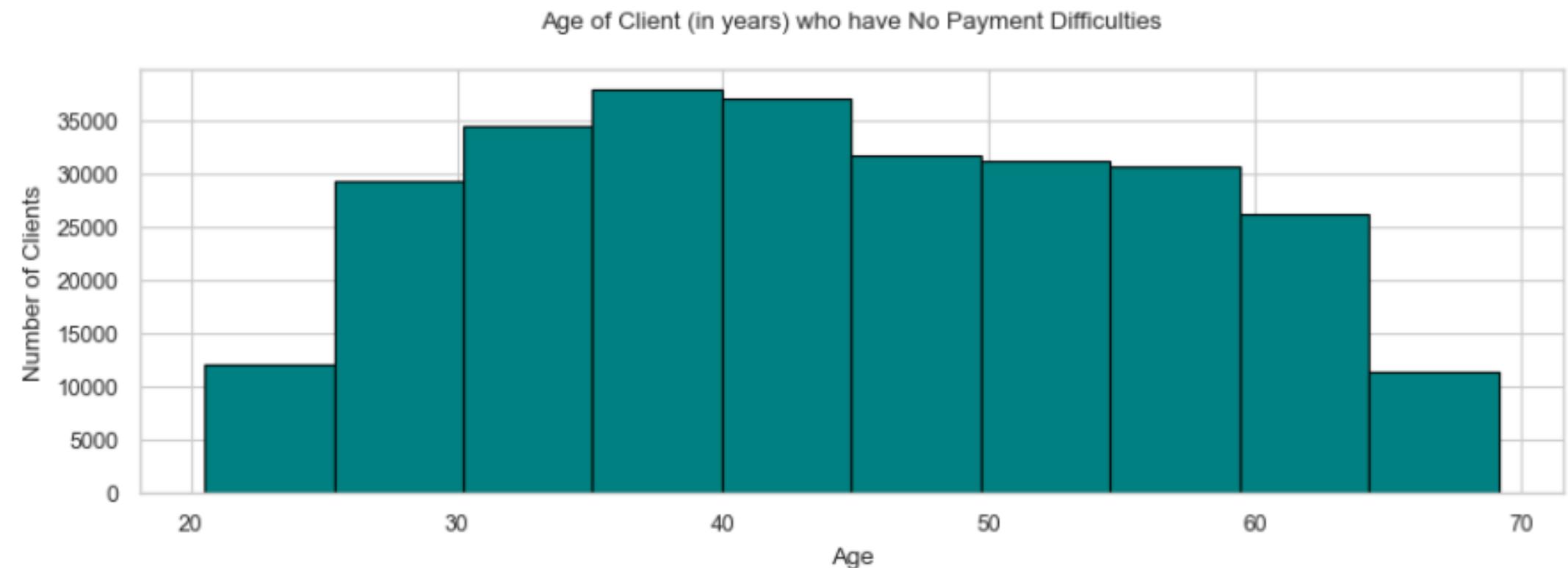
*There are some columns that are dropped and the rest are imputed*

### Label Encoding

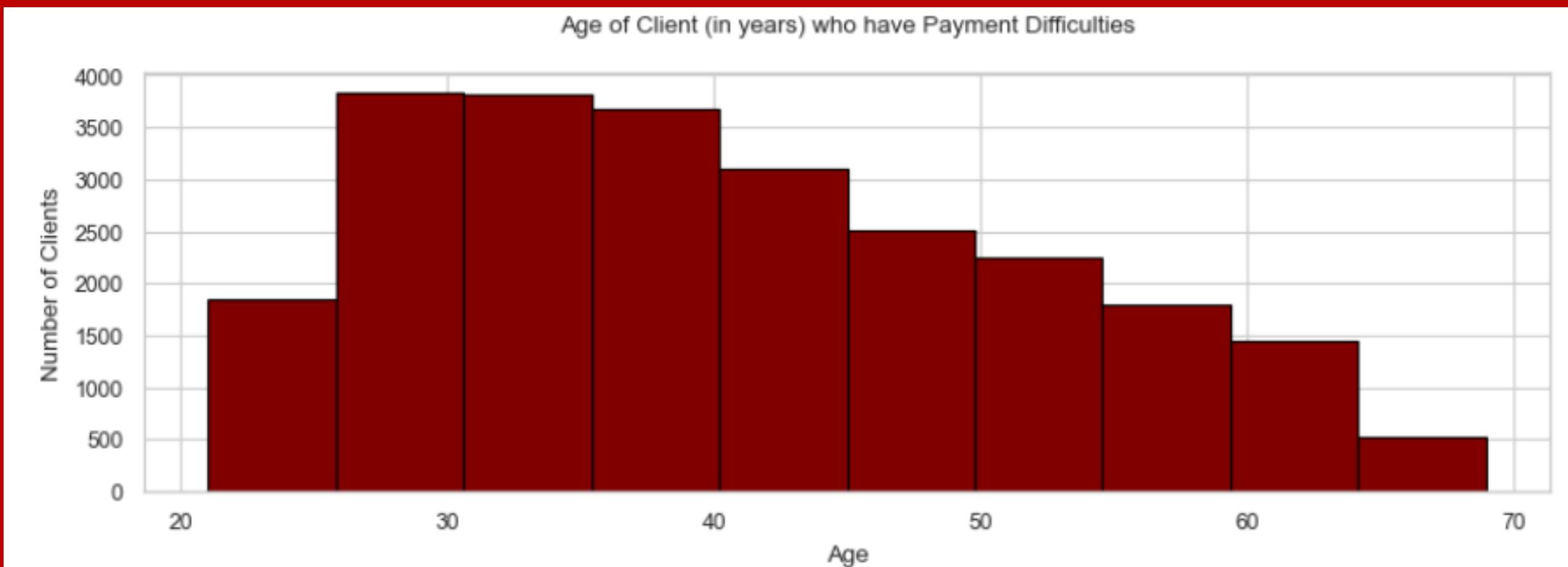
*Transform non-numerical to numerical labels*

## PREDICTION

*Predict client's repayment abilities with best machine learning model obtained before*

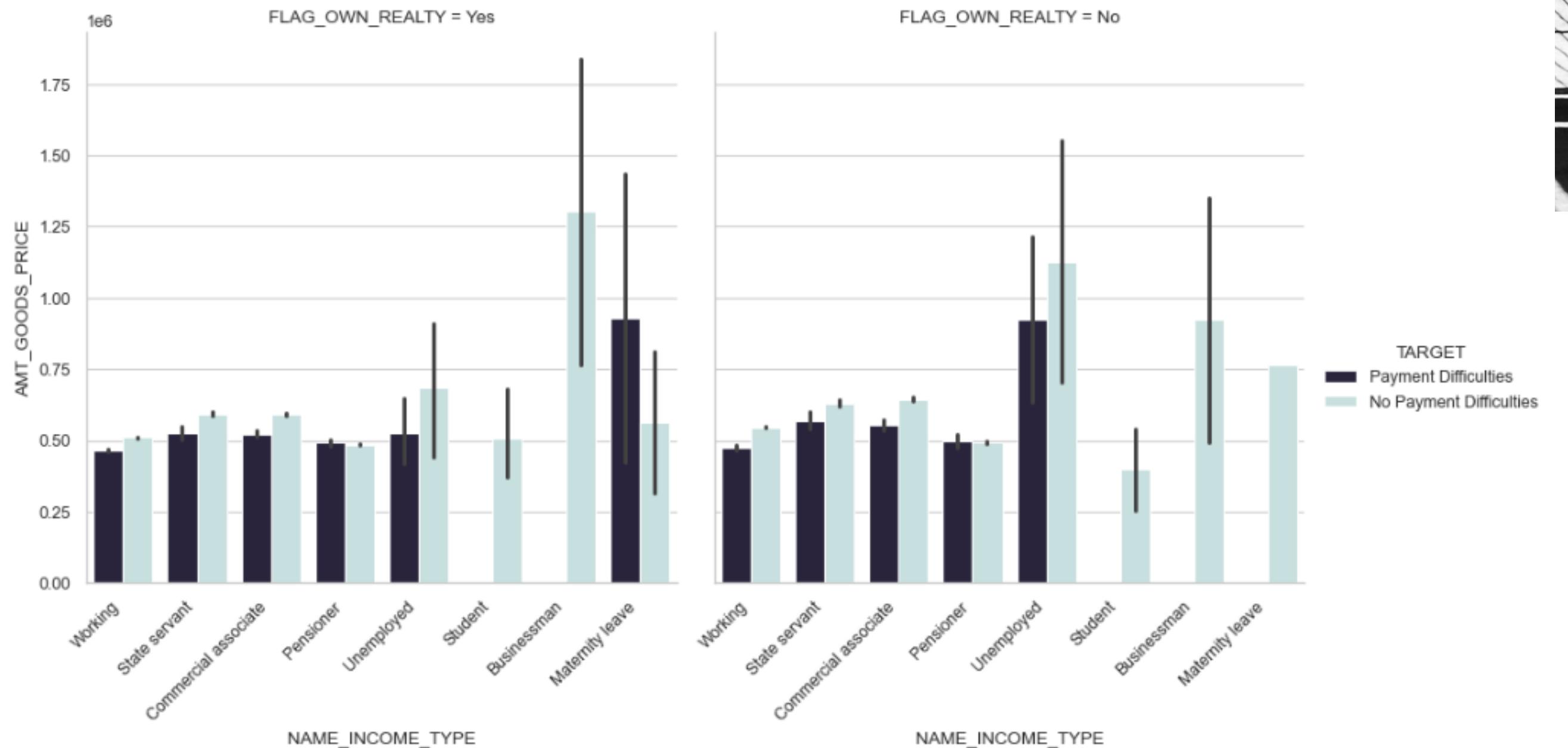


Most number of clients who apply for loans are in the range of 35-40 years. Meanwhile, the number of applicants for clients aged <25 or age >65 is very low.



- Clients who have **no payment difficulties** are clients in the range of **35-45** years. You can target these clients as your priority.
- While clients who have **payment difficulties** are client the range of **25-35** years.

C:\Users\yudhamulyana\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight  
self.\_figure.tight\_layout(\*args, \*\*kwargs)



All student clients have no difficulty repaying the loans whether with cash loan or revolving loan for a low to medium credit amount of the loan.

For the income type of **maternity leave** with **cash loans**, all the clients **have problems repaying the loans** for a **medium credit amount of the loan**. While all clients with maternity leaves and revolving loans have no difficulty repaying the loans.

For **unemployed** clients with **cash loans**, **more than 50%** of clients **have problems repaying loans** with **medium credit amounts of the loan**. While all unemployed clients with revolving loans have no difficulty repaying the loan.

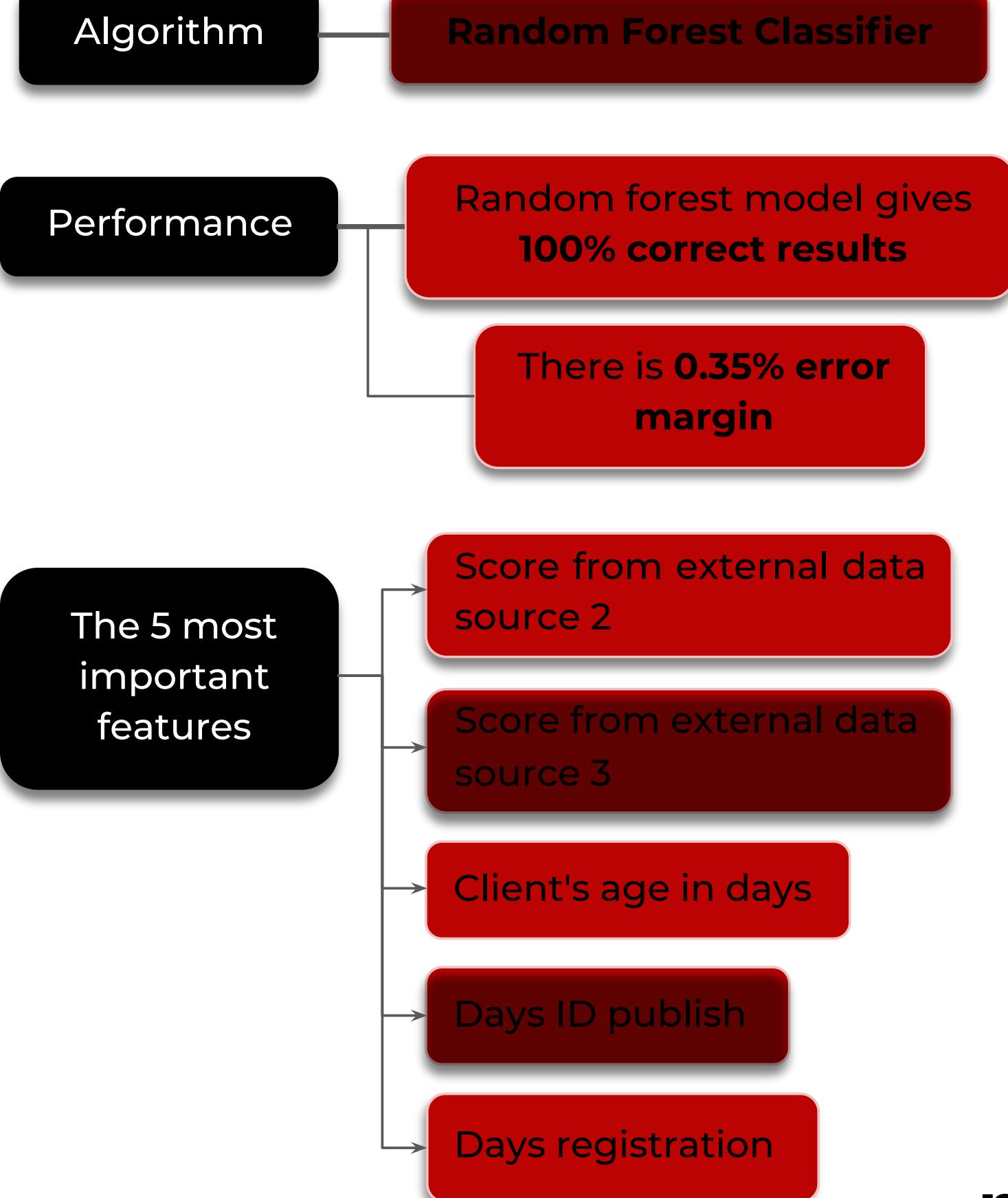
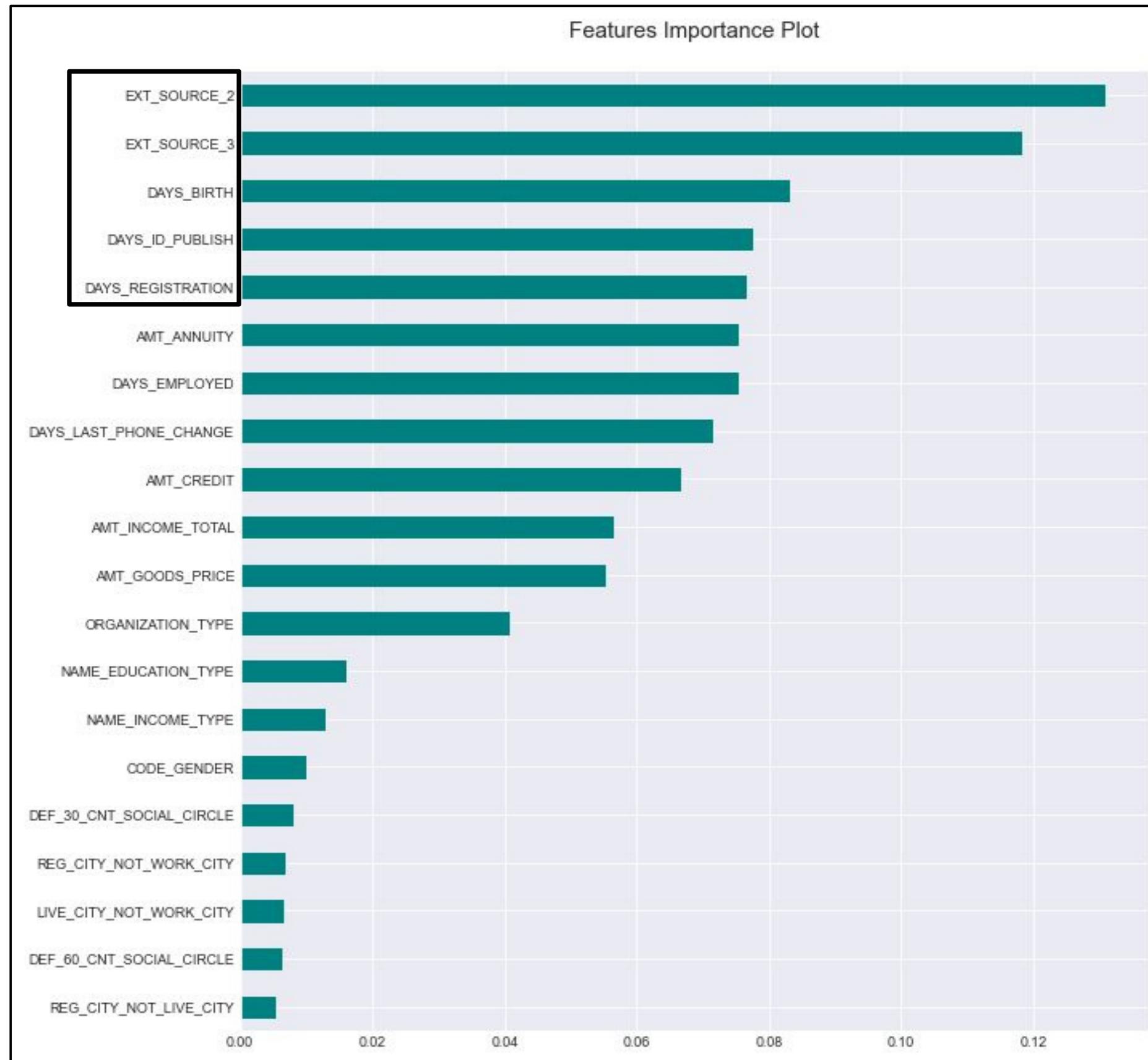
| Algorithm            | Training Accuracy Score | Testing Accuracy Score | Error Margin | ROC Score |
|----------------------|-------------------------|------------------------|--------------|-----------|
| Logistic Regression  | 67.16%                  | 67.29%                 | 0.13%        | 0.672     |
| Gaussian Naive Bayes | 60.24%                  | 60.39%                 | 0.15%        | 8         |
| Decision Tree        | 100%                    | 83.9%                  | 11.74        | 0.604     |
| Random Forest        | 100%                    | 99.65%                 | %            | 0.882     |
| K-Nearest Neighbor   | 91.56%                  | 88.07%                 | 0.35%        | 6         |
| Neural Network       | 70.01%                  | 69.48%                 | 3.79%        | 0.996     |

0.58%

5

The prediction accuracy of the train and test data in **Random Forest** model has a value that is not much different, it can be said that the model is very good, which is there is **no underfitting or overfitting**. So the **Random Forest** model was chosen as the **best model to predict client's repayment abilities**.

# BEST MODEL



# RECOMMENDATION

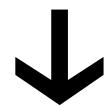
1. A client with an income type of **student** can be said to be a client who is **capable of repaying the loans** whether with a cash loan or revolving loan (100% of applications approved). But there only 0.005% of applications come from the student.
2. A client who works as an **accountant** can be said to be a client who is **capable of repaying the loans** (95% of applications approved). But, there is only 3.19% of applications come from an accountant. So do, the client who work as **high skill tech staff** and **manager**, they are capable of repaying the loans, but there are only a few applications that come from them



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

# RECOMMENDATION

1. Clients with **maternity leaves** and **cash loans** can be said to be a client who is **incapable of repaying the loan** (100% of applications rejected). On the contrary, all clients with maternity leave but taking revolving loans to have their applications approved.
2. For **unemployed** clients, more than 50% of them **have a problem repaying their loans** if they take **cash loan** contracts. Meanwhile, all unemployed client who takes revolving loans is capable of repaying the loan.



**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

- There are several features that have the most influence on the target:
- The more you own a car, the greater the chance of smooth payments
- The more active the credit, the greater the chance of smooth payment
- The higher the price of the item, the greater the chance of smooth payment
- The bigger the house, the greater the chance of smooth payments
- The higher the position or job, the greater the chance of smooth payment
- The longer you work, the greater the chance of smooth payment
- The younger you are when you apply for a loan, the less smooth it is
- The higher the city rating number, the less smooth it is
- Applicants who have recently changed their telephone number are more likely to fail to pay

- There are several features that have the most influence on the target:
- The more you own a car, the greater the chance of smooth payments
- The more active the credit, the greater the chance of smooth payment
- The higher the price of the item, the greater the chance of smooth payment
- The bigger the house, the greater the chance of smooth payments
- The higher the position or job, the greater the chance of smooth payment
- The longer you work, the greater the chance of smooth payment
- The younger you are when you apply for a loan, the less smooth it is
- The higher the city rating number, the less smooth it is
- Applicants who have recently changed their telephone number are more likely to fail to pay

# LINK / PORTFOLIO



**GITHUB**



**LINKEDIN**

