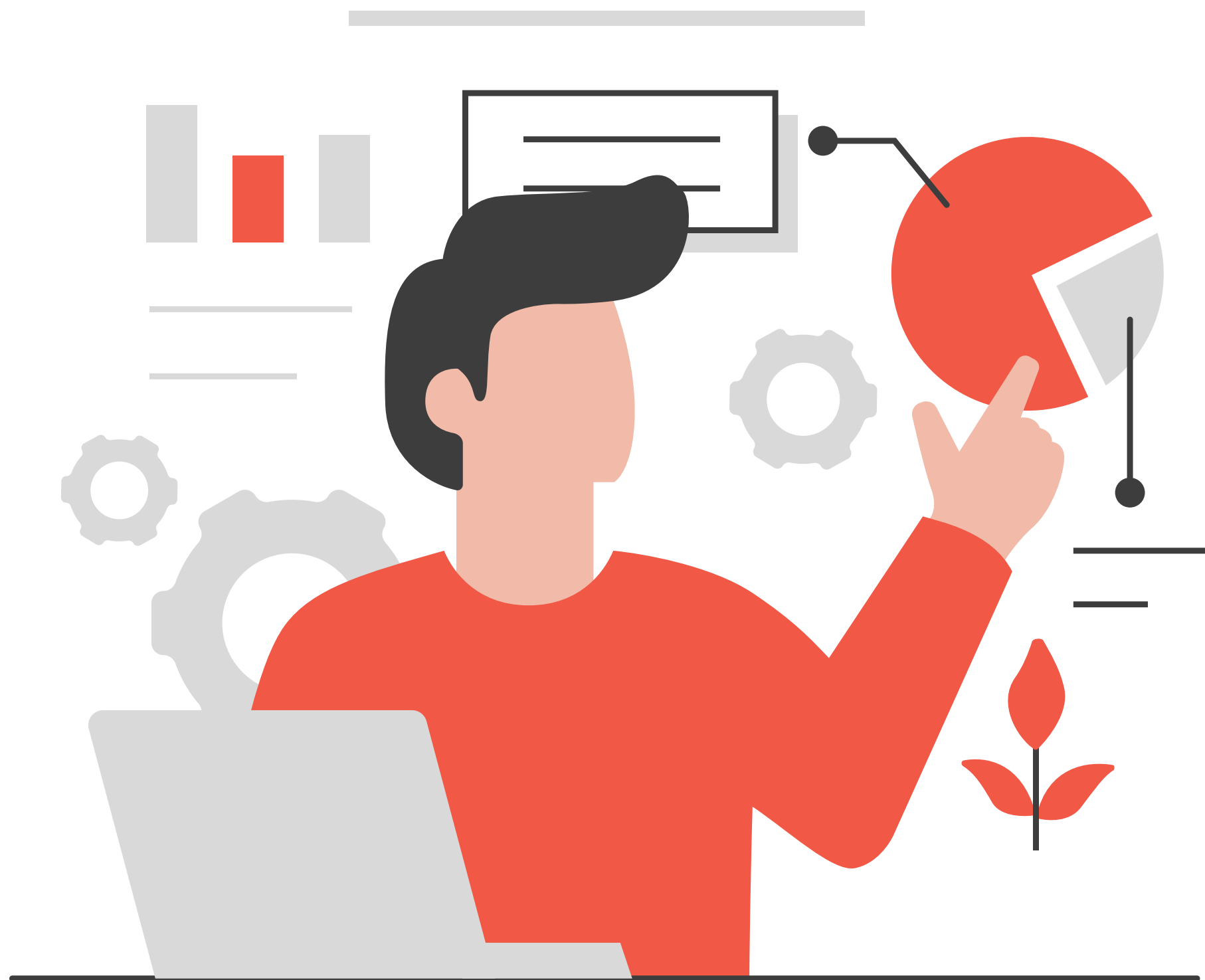


HOME CREDIT SCORECARD MODEL

By Akbar Nur Rizqi





WORKFLOW

01. Problem Research

02. Data Pre-Processing

03. Data Visualization and
Business Insight

04. Machine Learning
Implementation and Evaluation

05. Business Recommendation

PROBLEM RESEARCH



PROBLEM

How to measure creditworthiness of the clients?

GOAL

Identify and predict the characteristics of customers who will experience difficulties and no difficulties in repaying loans.

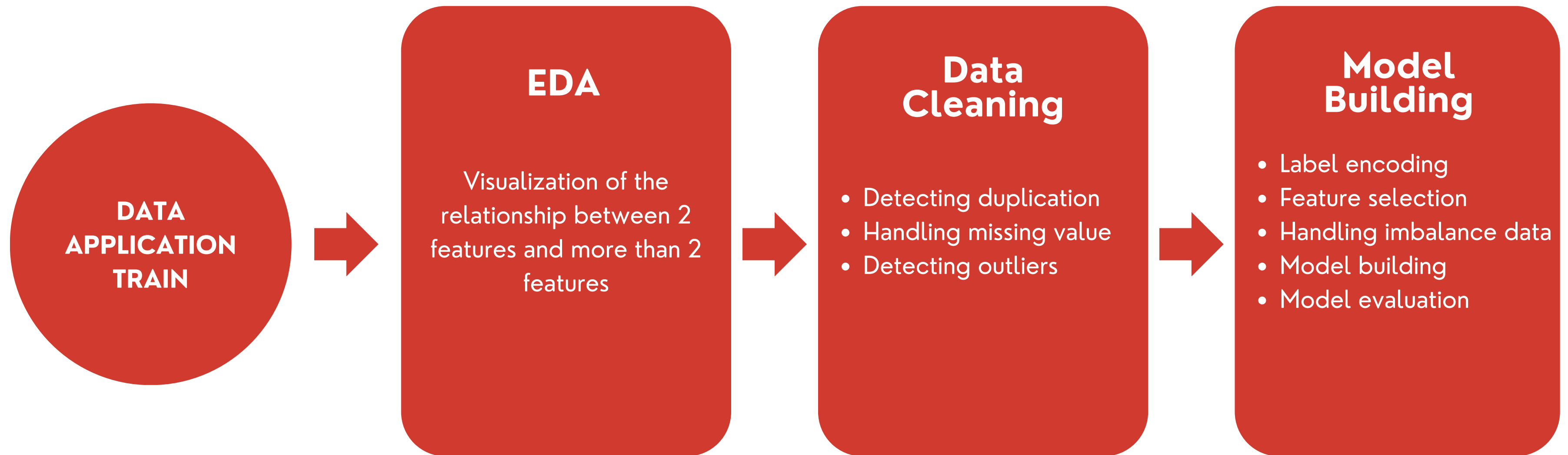
OBJECTIVE

Creating a credit scoring model to manage, understand, and model credit risk that will be handled optimally.

BUSINESS METRICS

Can be used to decide whether a client's credit application is accepted or rejected

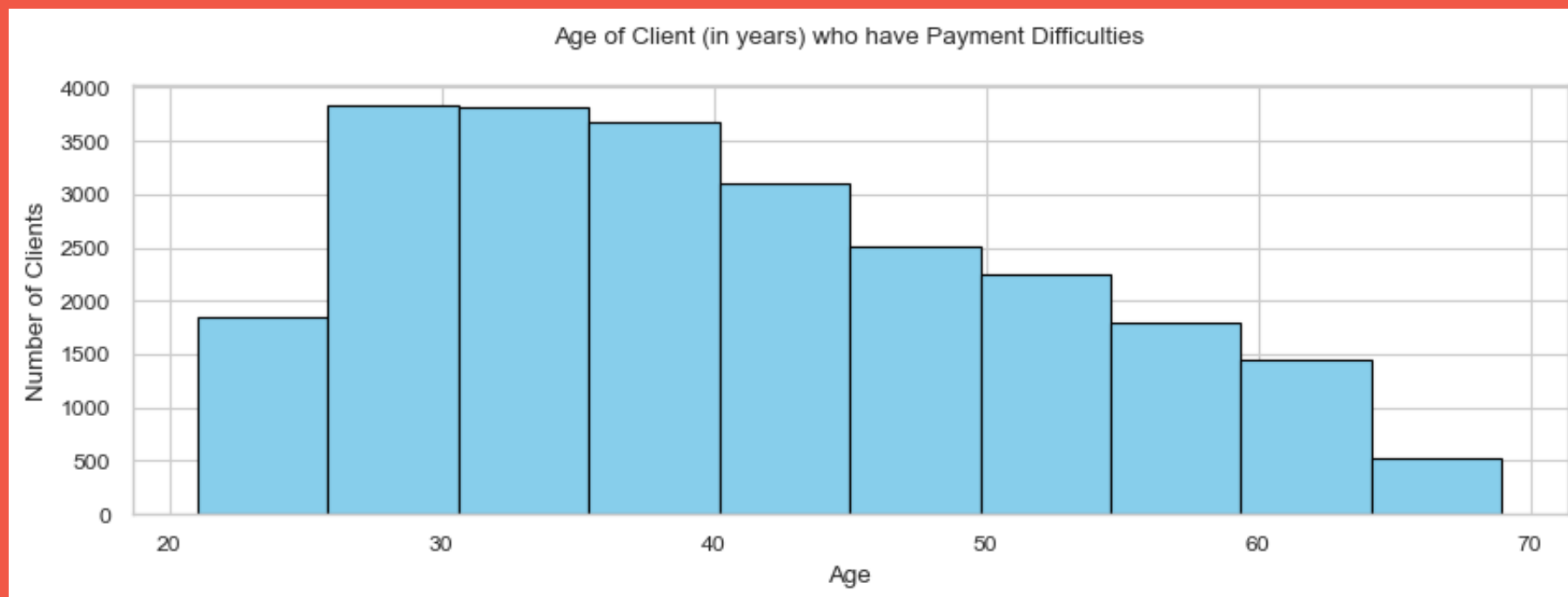
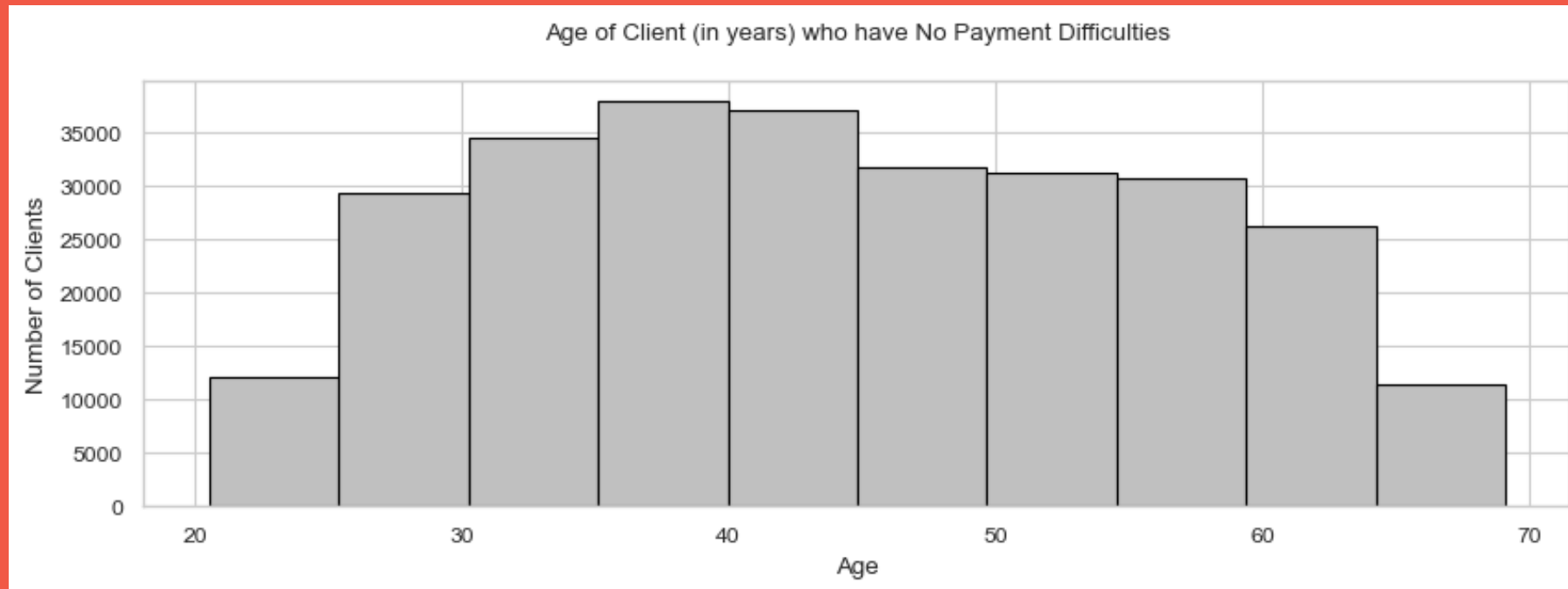
DATA PRE-PROCESSING



DATA PRE-PROCESSING



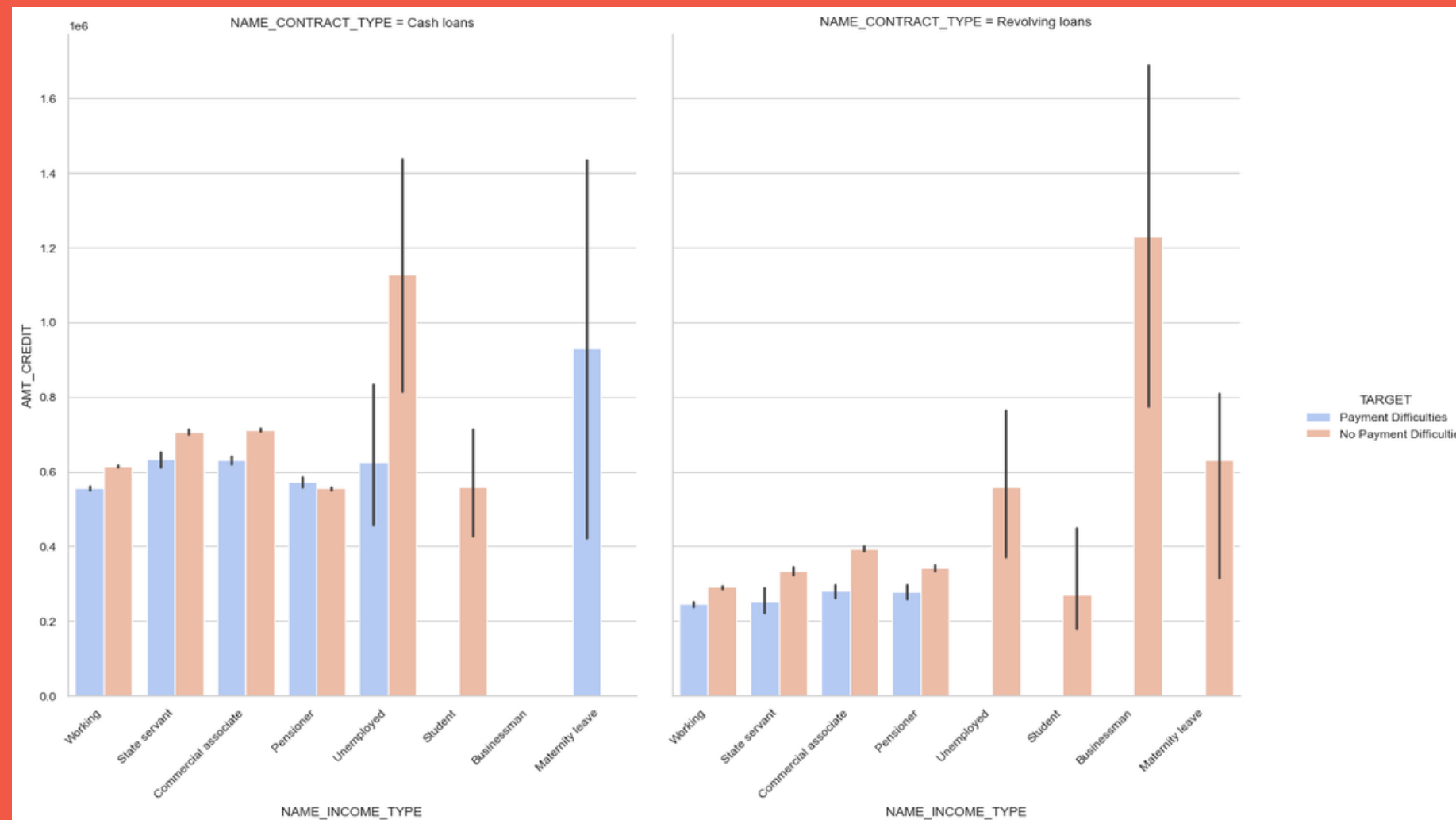
DATA VISUALIZATION AND BUSINESS INSIGHT



- Most number of clients who apply for loans are in the range of 35-40 years and the number of applicants for clients aged <25 or age >65 is very low.
- Clients who have no payment difficulties are client with an age range of 35-45 years. Meanwhile, clients who have payment difficulties are clients with an age range of 25-35 years.



DATA VISUALIZATION AND BUSINESS INSIGHT



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



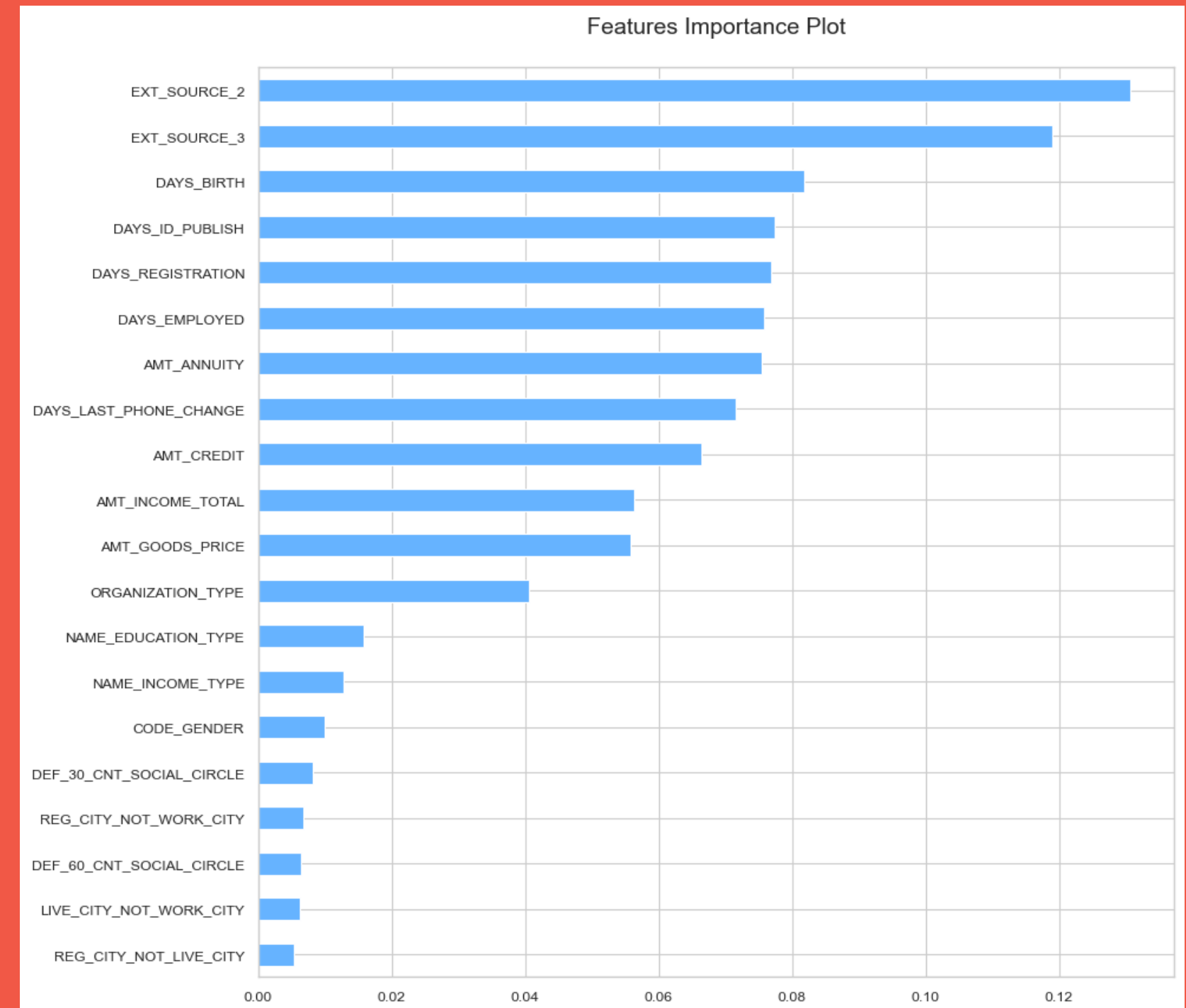
MACHINE LEARNING IMPLEMENTATION AND EVALUATION

Best Model

Model Comparison

Models	Training Accuracy Score	Testing Accuracy Score	ROC Score
Logistic Regression	67.16%	67.29%	0.6728
Gaussian Naive Bayes	60.24%	60.39%	0.604
Decision Tree	100%	88.3%	0.8826
Random Forest	100%	99.65%	0.9965

- **Algorithm:** 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
- **Performance:** Correct result of **Random Forest** model is 100%



BUSINESS RECOMMENDATION

Business Insight

- Clients who have no payment difficulties are client with an age range of 35-45 years. Meanwhile, clients who have payment difficulties are clients with an age range of 25-35 years.
- 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 clients with type of maternity leave and unemployed customers with cash loans have problems repaying loans with moderate loan amounts. As for revolving loans, there is no difficulty in repaying the loan
- The highest percentage of clients with problems in repaying loans are clients with Low-Skilled Worker jobs at 17% and clients who works as an accountant, High skill tech staff, Managers can be said to be clients who are capable of repaying the loans



Business Recommendation

- Creating campaigns for students, accountants, high skill tech staff, managers, and clients with age ranges of 35-45 years to be interested in applying for a loan
- Need further analysis, you can do a survey (such as Family Status, House/Flat Type and Ownership Status, The Number of Children, Occupation) to find out if there are any problems when taking a cash loan contract, especially clients on maternity leave or unemployed. So, going forward, if you have a client with that type of income, you can recommend the right type of contract for their application to be approved.

THANKS FOR WATCHING

My Github

<https://github.com/nathing101/ScoreCard-Model>

