Credit Scoring with Machine Learning

By Chandra Ayunda Apta Soemedhy

Project-Based Virtual Intern : Data Scientist -Home Credit Indonesia x Rakamin Academy

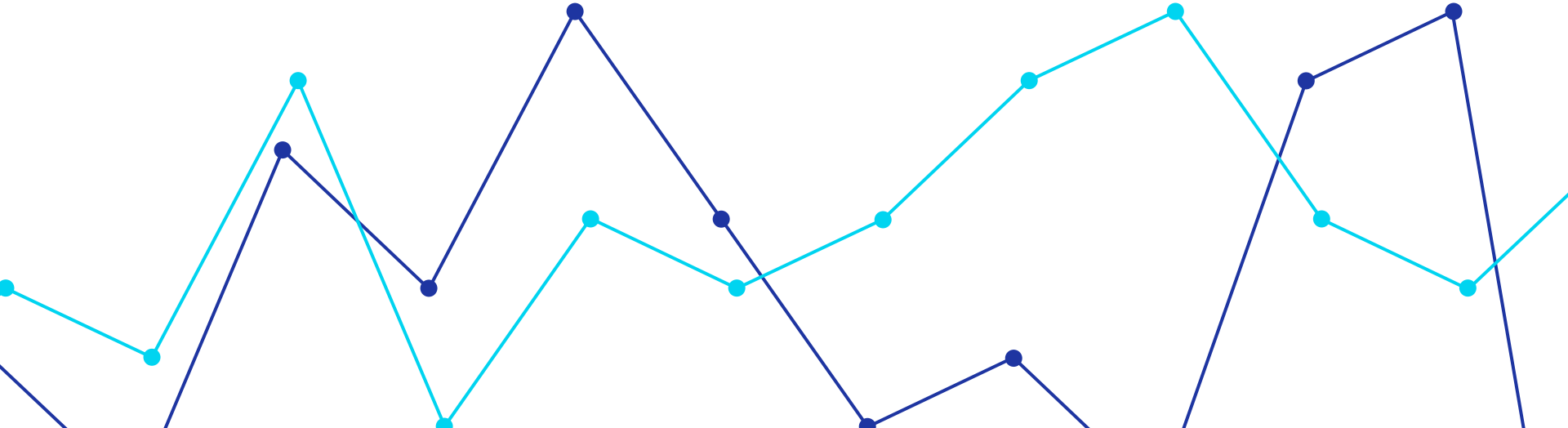


Table of Contents

Problem Research

Background

Home Credit is currently using various statistical methods and Machine Learning to make credit score predictions. With the available data, we can maximize business potential by ensuring that customers who are able to make payments are not rejected when applying for a loan, and that loans are provided with a principal, maturity and repayment calendar that will motivate customers to succeed.

Action

1. Preprocessing Data
2. Visualization and Business Insight
3. Modelling and Evaluation
4. Predict customer repayment with application test data
5. provide business recommendations to companies

Data Preprocessing

Prepare Data

- 122 columns
- 307.511 row

EDA

Multivariate Analysis

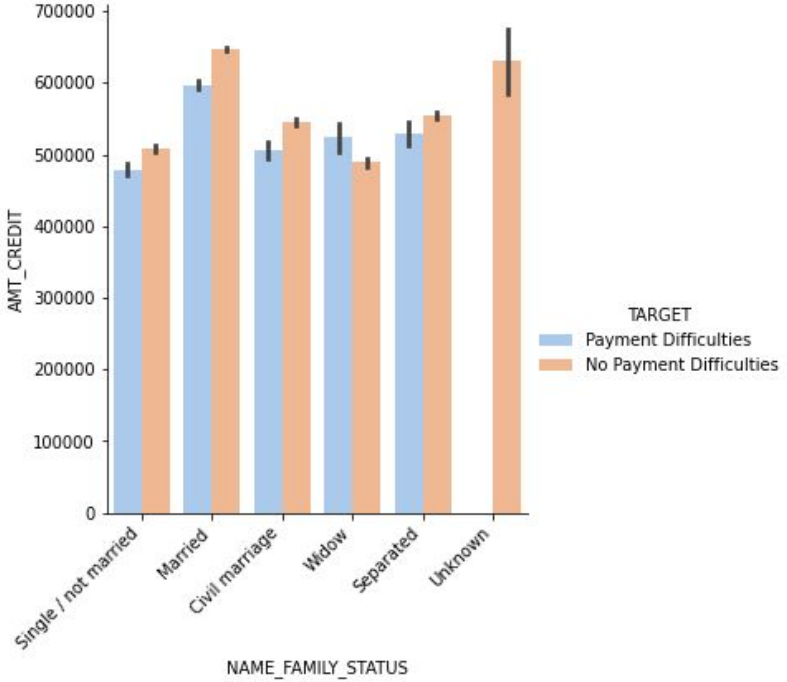
Data Cleaning

- Check and handling missing value
- Label Encoder
- SelectKbest(Feature Selection)
- Handling Imbalance Data

Modeling and Evaluation

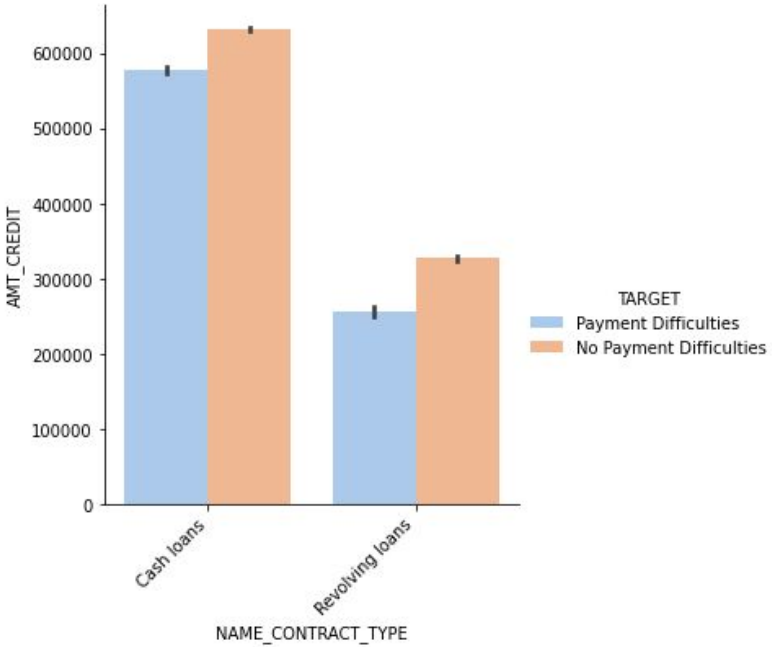
- Modeling and model selection
- Model evaluation
- Data predict

Multivariate Analysis

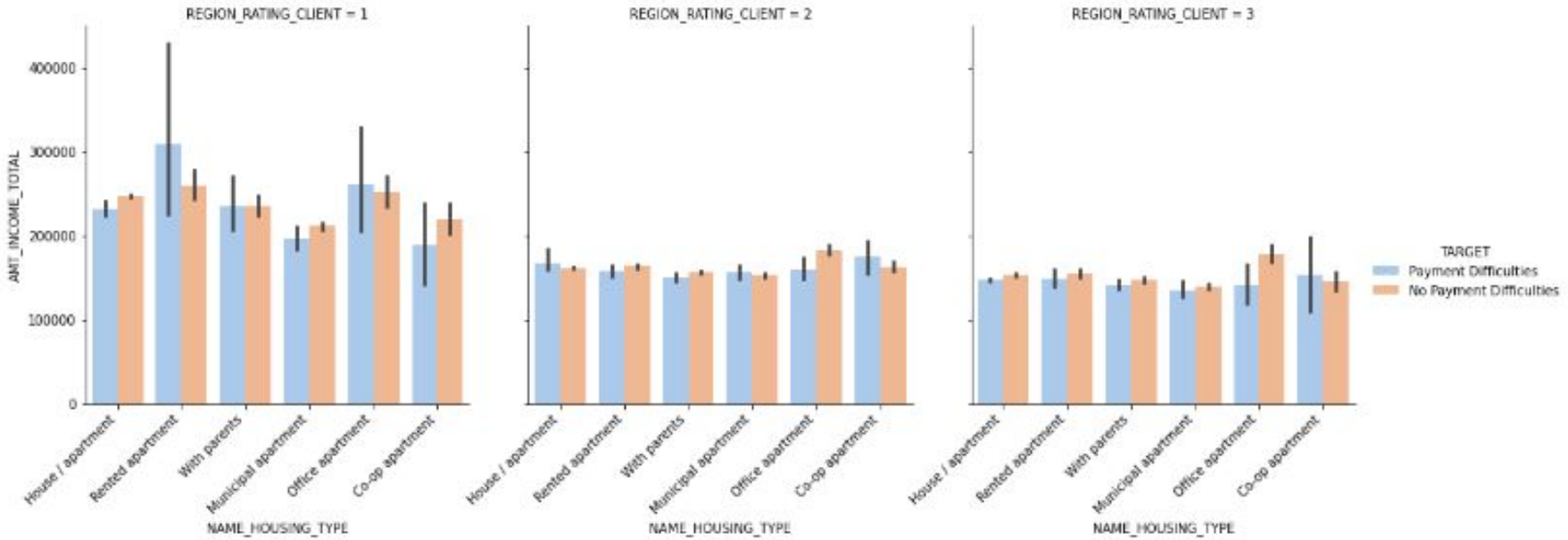


Customers with the largest amount of credit and experiencing payment difficulties are married customers.

The majority of customers use the cash loans system with the highest level of difficulty and ease of payment.



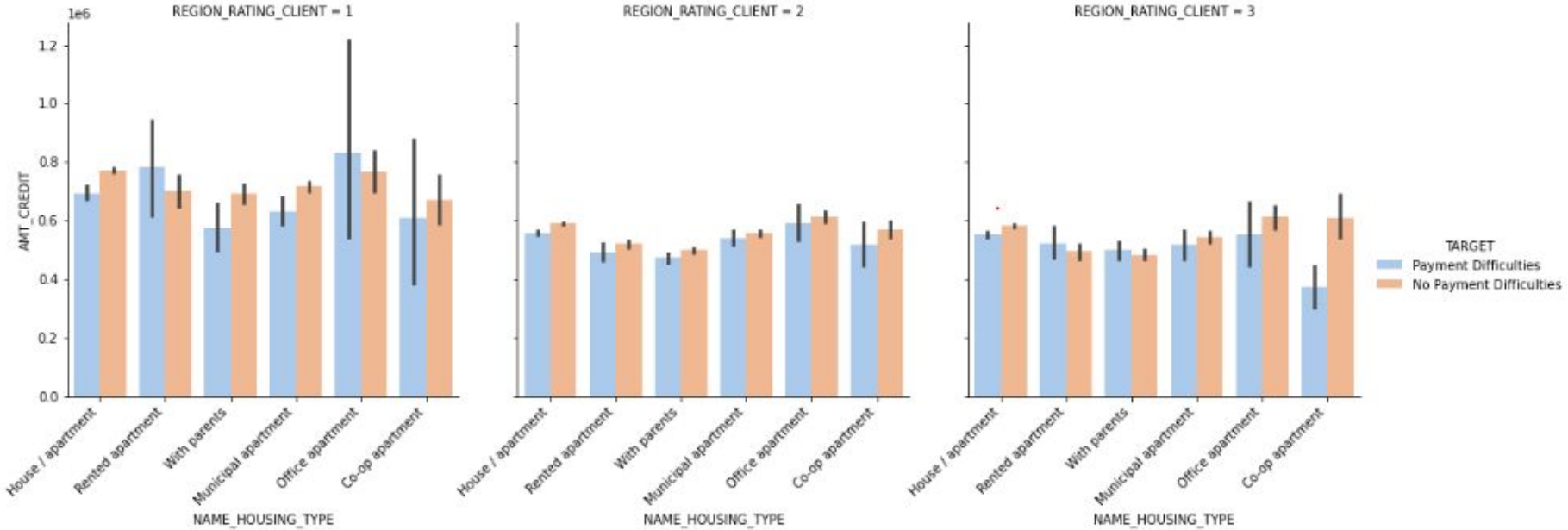
Multivariate Analysis



- The customer with the highest income is in region_client_rating = 1 with a high payment rate too.

- Customers with the highest level of payment difficulty have the rented apartment and office apartment type of house.

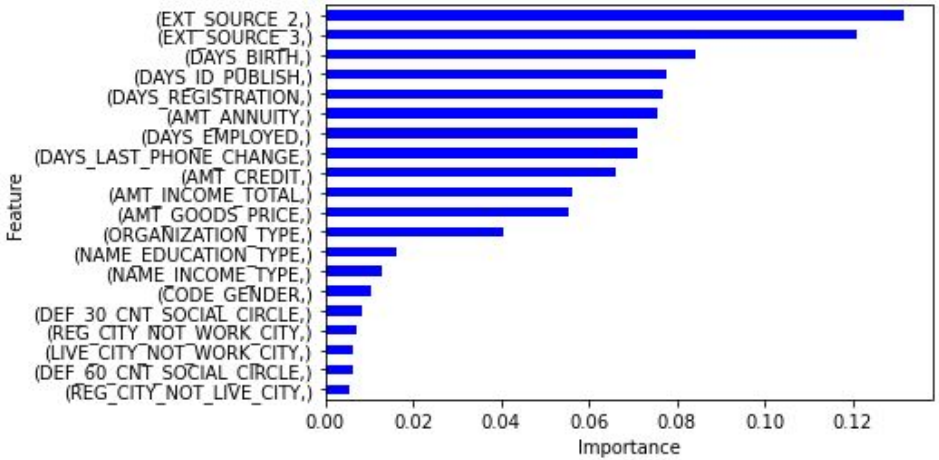
Multivariate Analysis



- The customer with the highest number of credit in region_client_rating = 1.

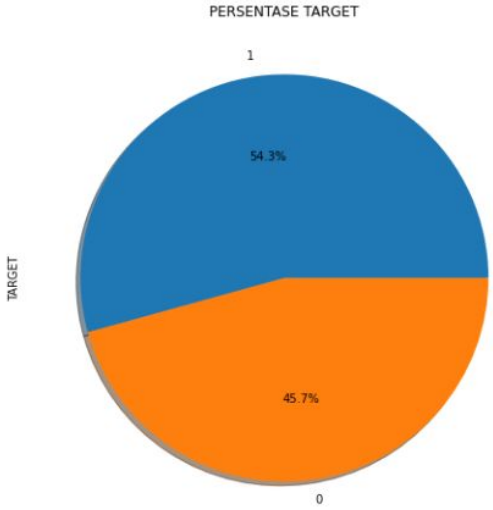
- Customers with the highest level of payment difficulty have the rented apartment and office apartment type of house.

Modelling and Evaluation



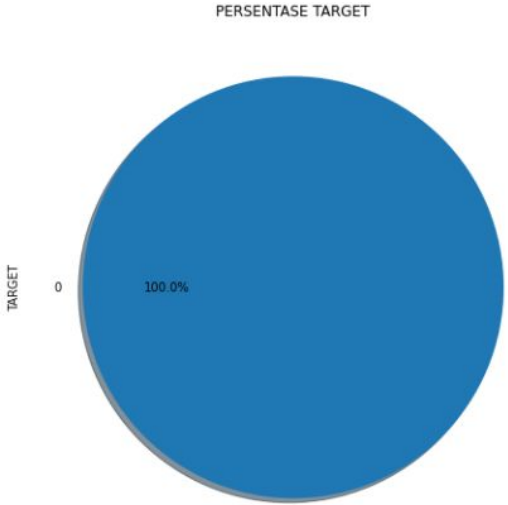
	Logistic Regression	Random Forest	Naive Bayes
Accuracy	67%	99%	61%
Precision (0)	67%	100%	57%
Precision (1)	68%	99%	69%
Recall (0)	69%	99%	83%
Recall (1)	66%	100%	38%

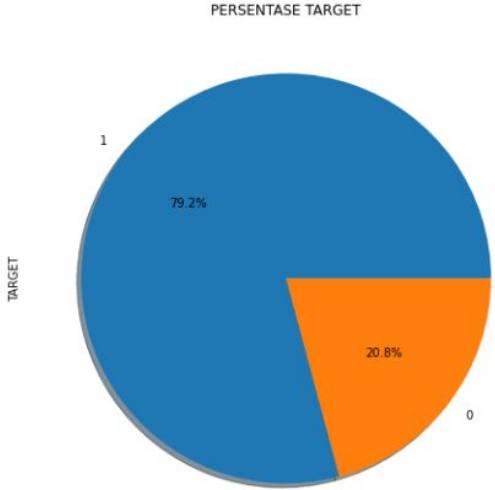
Modelling and Evaluation



Logistic Regression

Random Forest





Naive Bayes

Business Recomendation

1. More focus on 20 factors which are the most important variables in the data for the customer credit scoring process.
2. focuses more on offering cash loan contract types by prioritizing region_rating_client = 1.
3. Focusing on customers with incomes above 200,000 with married status and rented apartment and office apartment types.
4. Maintain customer trust by establishing good relationships.