



HOME CREDIT SCORECARD MODEL

Ryan Rachmad Hidayat

Read more



Table of Content

- 
- 01 Project Backgorund
 - 02 Problem Research,
Purpose, and Dataset
 - 03 Data Pre-Processing
 - 04 Data Visualization and
Business Insight
 - 05 Machine Learning
Implementation and Evaluation
 - 06 Business Recommendation

Project Background

PT Home Credit Indonesia is a technology-based financing company with a wide range of store partners in more than 200 cities in Indonesia. PT Home Credit Indonesia has been operating since 2013, Home Credit has now grown to become a trusted financial partner for millions of customers. PT Home Credit Indonesia also encourages open access to transparent financial services with a fast process and is able to help people meet their needs in a planned manner including managing their finances and installments properly.

In providing a loan to someone, PT Home Credit Indonesia will definitely assess the credit risk of the borrower. To assist in the credit risk assessment process, PT Home Credit Indonesia needs an accurate and reliable credit score model. A credit score is a metric used to predict the likelihood that a borrower will default within a certain period of time. This score is calculated based on various factors such as credit history, income, employment, and other demographic information.



Problem Statement

The main objective of this project is to build an accurate and reliable credit score model for PT Home Credit Indonesia. This model will be used to evaluate potential borrowers and help make better lending decisions.

Purpose

By solving this problem, Home Credit was able to improve their credit risk assessment process, reduce default losses, and ultimately improve the profitability and sustainability of their business.

[Read more](#)

page 04/10

Dataset

The data used for this analysis was provided by the Home Credit Indonesia team. There are two datasets used, the Train dataset and the Test dataset. Both datasets consist of around 307,000 rows that contain data on customers who are delinquent and non-delinquent when repaying the loans that have been provided by Home Credit Indonesia. The difference between the two files is the train data (with TARGET) and the test data (without TARGET) with 122 columns and 121 columns respectively.

[Read more](#)

page 04/10

Data Perprocessing

The process of performing data pre-processing is:

- Handling missing values
- Handling duplicated data
- Handling Outlier
- Feature Transformation (One hot encoding and label encoding)
- Split Data
- Handling Imbalance Data With Oversampling and Undersampling

[Read more](#)

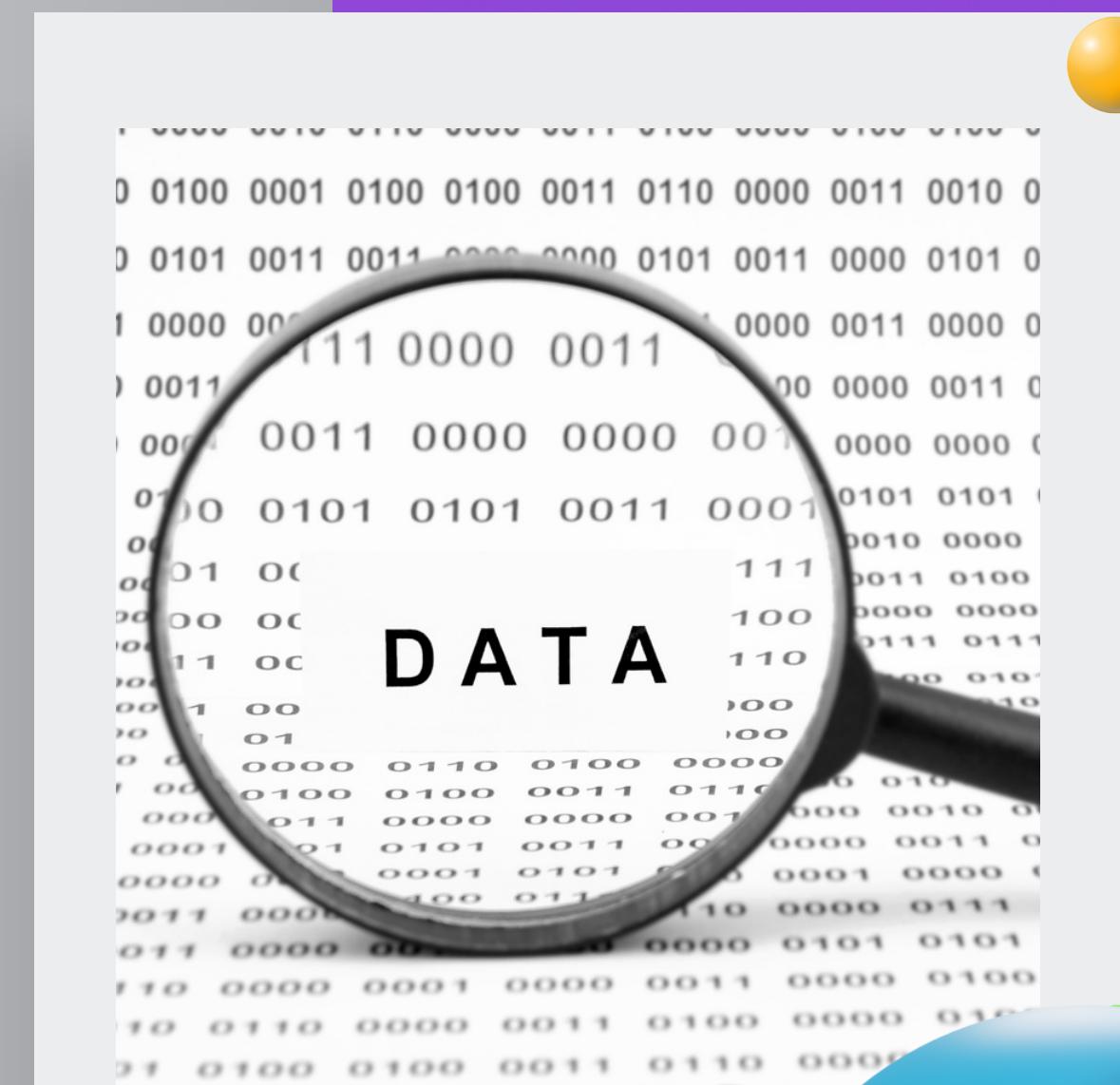


121

Column Before
PreProcessing

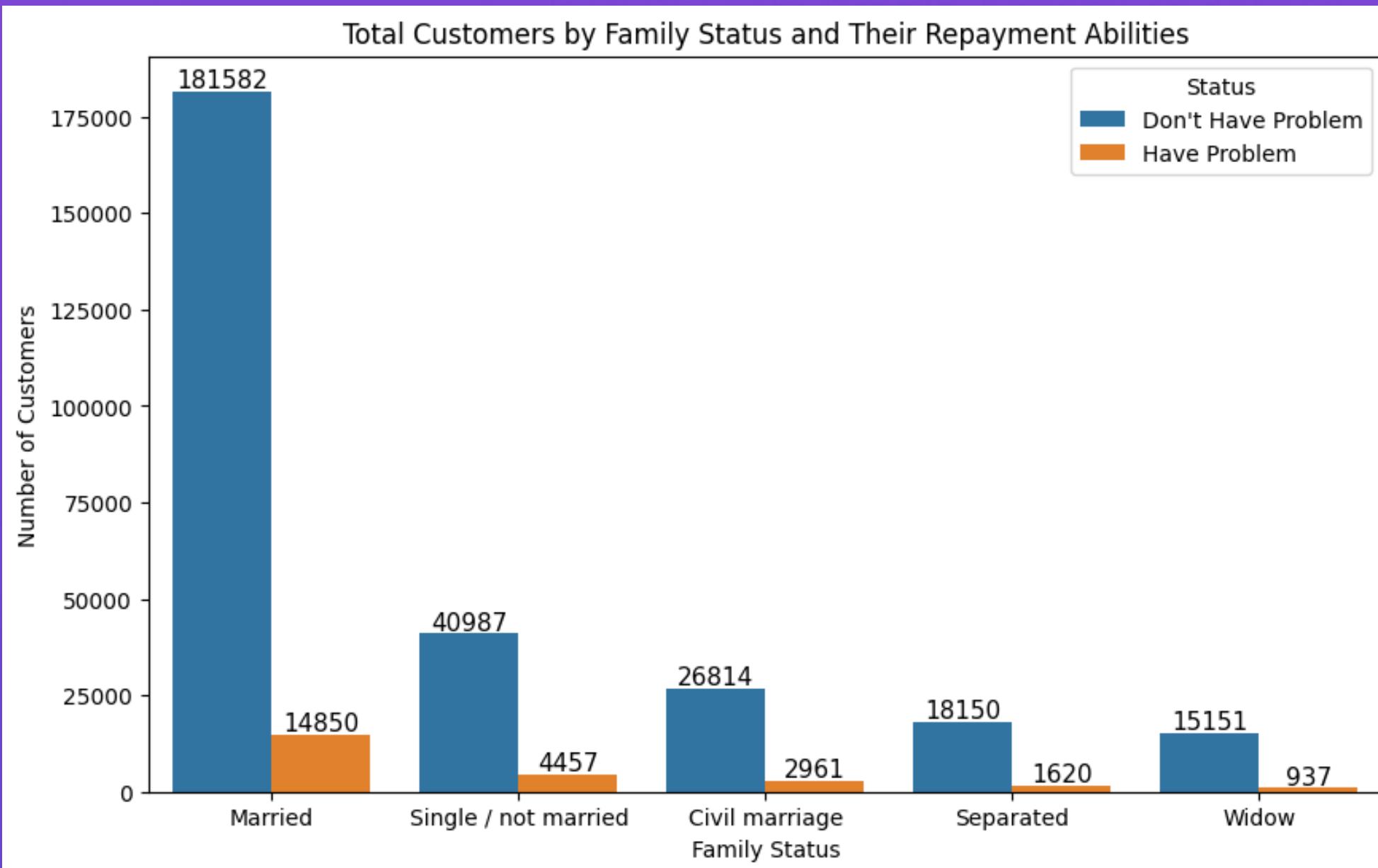
97

Column After
PreProcessing



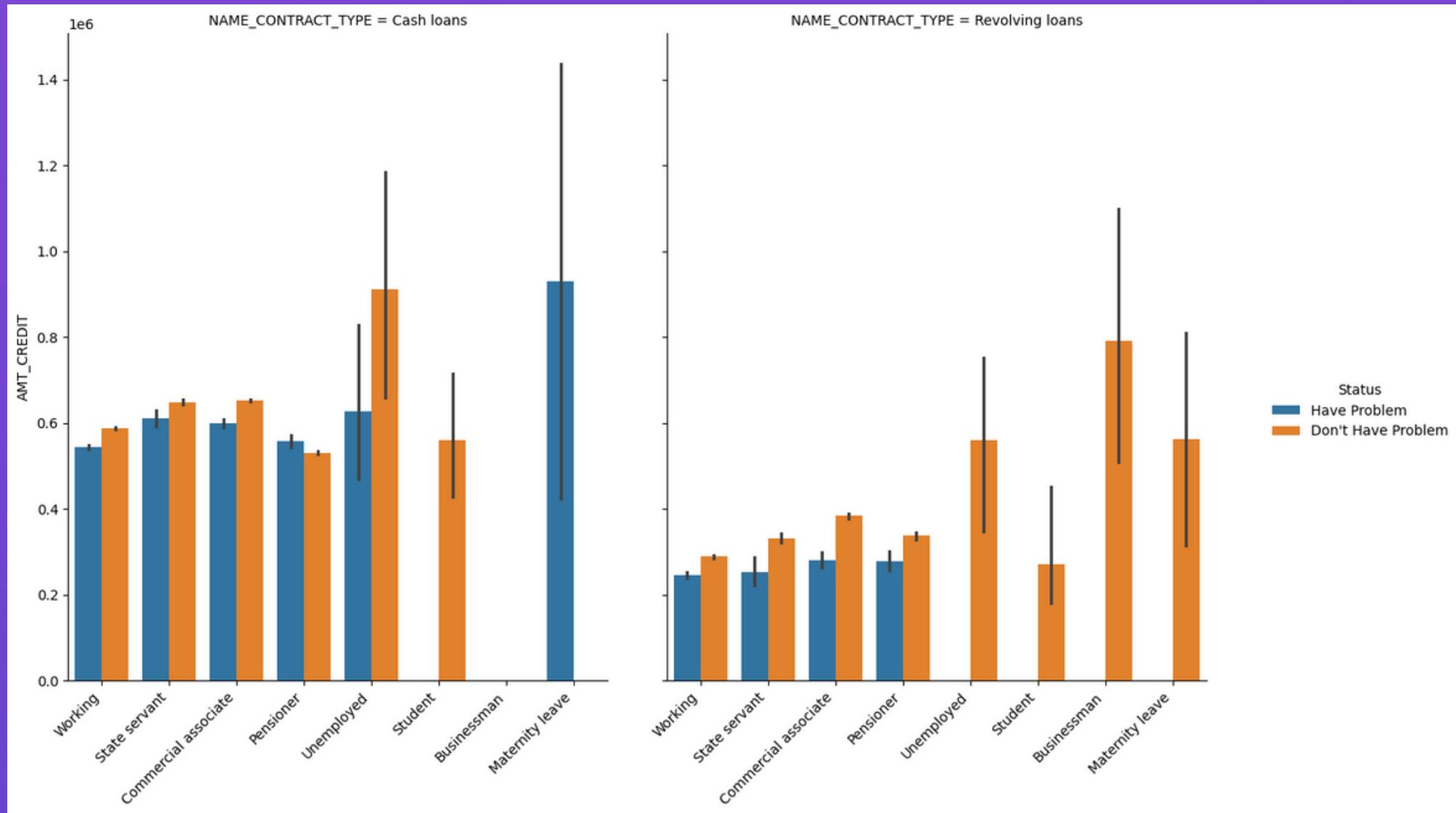
Business Insight

Married people seem to be a big target market for credit products. While most married customers are able to repay their credit, there is a small segment that experiences payment problems.



Business Insight

All student clients had no difficulty in repaying their loans, either with cash loans or revolving loans. In addition, people on maternity leave had no difficulty repaying revolving loans, but had difficulty repaying cash loans. All businessman opted for the revolving loan and none of them experienced any difficulty in repaying the loan.





Machine Learning Model

The Machine Learning model used in this analysis:

- Logistic Regression (Use Oversampling and Undersampling).
- XGBoost (Use Oversampling and Undersampling).
- Random Forest (Use Oversampling and Undersampling).

Read more

Model Comparison

Sampling	Model	Accuracy of Train Data	Accuracy of Test Data	Precision	Recall	F1-Score
Oversampling	Logistic Regression	58%	58%	58%	58%	58%
Oversampling	XGBoost	95%	95%	96%	95%	95%
Oversampling	Random Forest	100%	95%	100%	100%	100%
Undersampling	Logistic Regression	57%	57%	57%	57%	57%
Undersampling	XGBoost	81%	66%	81%	81%	81%
Undersampling	Random Forest	100%	66%	100%	100%	100%

The prediction accuracy of the training data and test data in the Random Forest model using oversampling has a value that is not much different. Although the XGBoost model has no difference in the accuracy of training data and test data, the interpretation of the random forest model is easier to explain. Therefore, this analysis uses the random forest model to predict the ability to pay customers.

Read more

Business Recommendation

1. Since there are about 92% of customers who do not have payment problems, we can provide reward points that can be exchanged for gifts or discounts to increase customer interest.
2. Customers who have payment problems are mostly married customers, so we can organize a seminar on financial planning or financial consultation.
3. We can do direct marketing or business exhibitions to reach more potential customers from businessmen and students.
4. Because it has good accuracy in both training and test data, we can use random forest to help the team determine whether a customer has a problem repaying credit or not.



Thank You Everyone

You can see the entire project documentation here!

<https://github.com/Ryan-Rachmad-Hidayat/Home-Credit-Score-Card-Model>

