HOME CREDIT Rakamin Academy

HOME CREDIT SCORECARD

Nabila Putri Listyanto

TABLE OF CONTENTS

1 Problem Research

2 Data Visualization & Business Insight

3 Data Preprocessing

4 ML Models & Model Evaluation

5 Business Recommendation

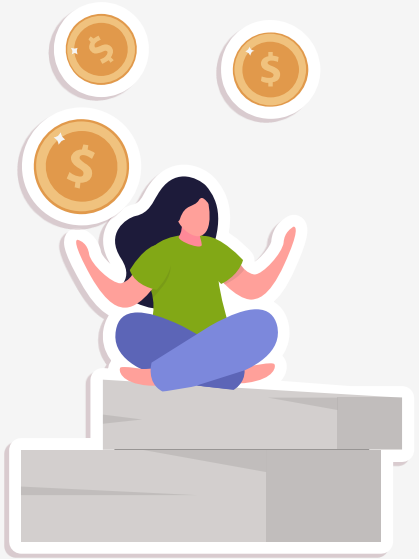
PROBLEM RESEARCH

Background

Companies ask to unlock the maximum potential of their data. Doing so will ensure that clients capable of repayment are accepted and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

Objective

Identify customers with characteristics of potential clients who have difficulty repaying loans.



Goal

Improve loan approval process by identifying potential clients who are likely to have difficulty repaying loans.

Metrics

Classification Report and AUC Score

FLOW

Exploratory Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Data Preprocessing

- Data Cleaning
- Data Imputation
- Feature Engineering
- Feature Selection
- Handling Imbalanced Data
- Data Splitting
- Feature Scaling

Models Building

Without Hyperparameter Tuning

- Logistic Regression
- Random Forest

With Hyperparameter Tuning

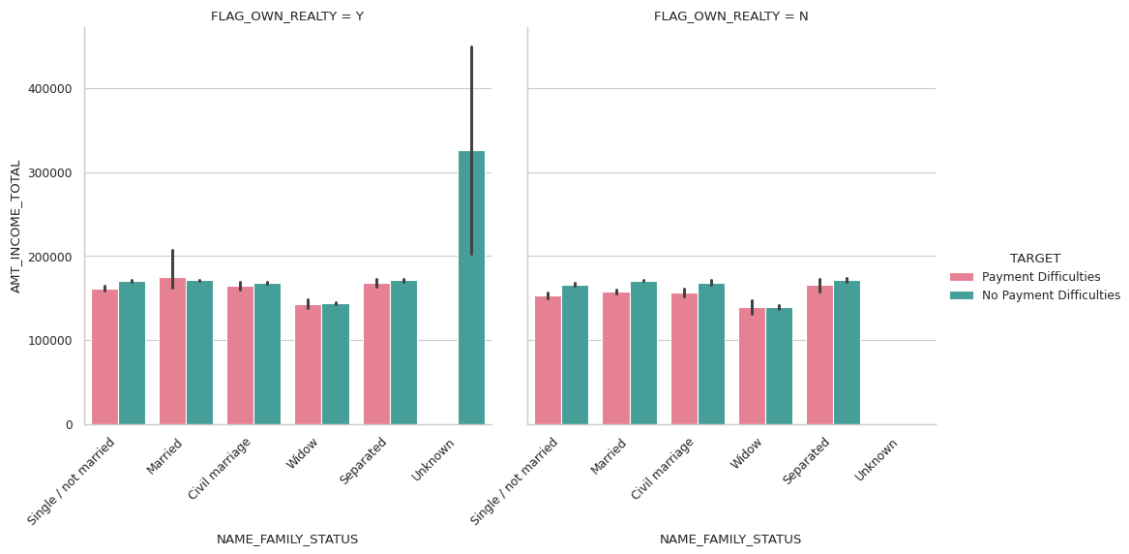
- Logistic Regression
- Random Forest

Model Evaluation

- Classification Report
- Confusion Matrix
- ROC Score

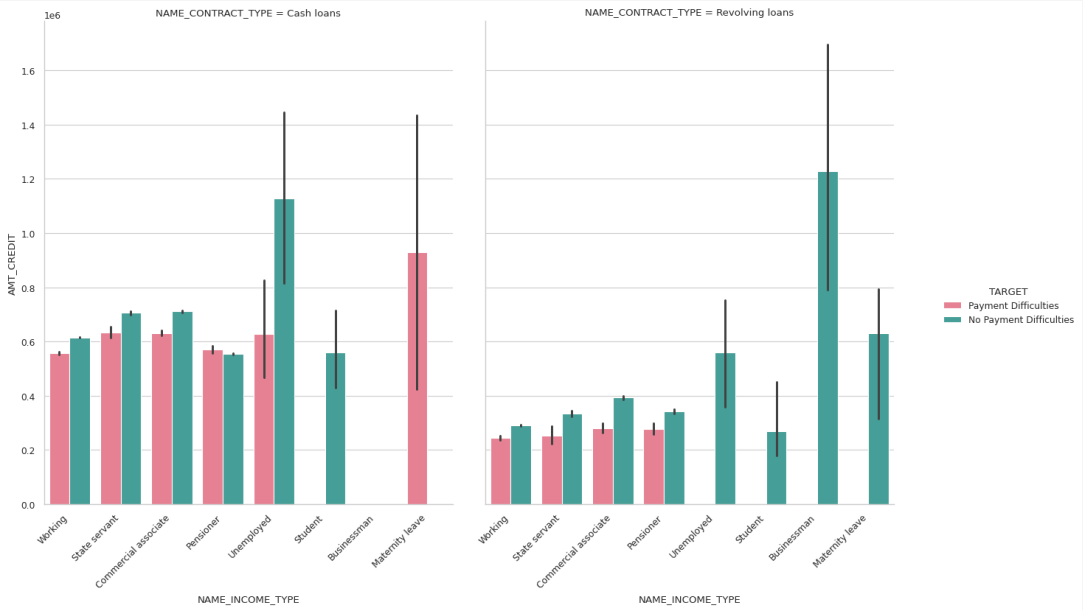
DATA VISUALIZATION & BUSINESS INSIGHT

Clients who are married and own a house/flat struggle with repaying loans for a medium amount of income, whereas those who do not own a house/flat do not face the same problem.



DATA VISUALIZATION & BUSINESS INSIGHT

Clients who are on maternity leave and have cash loans struggle with repaying loans for medium credit amounts, whereas those on maternity leave with revolving loans have no repayment difficulties. More than half of unemployed clients with cash loans face problems repaying medium credit amounts, but those with revolving loans have no issues..



MACHINE LEARNING MODELS

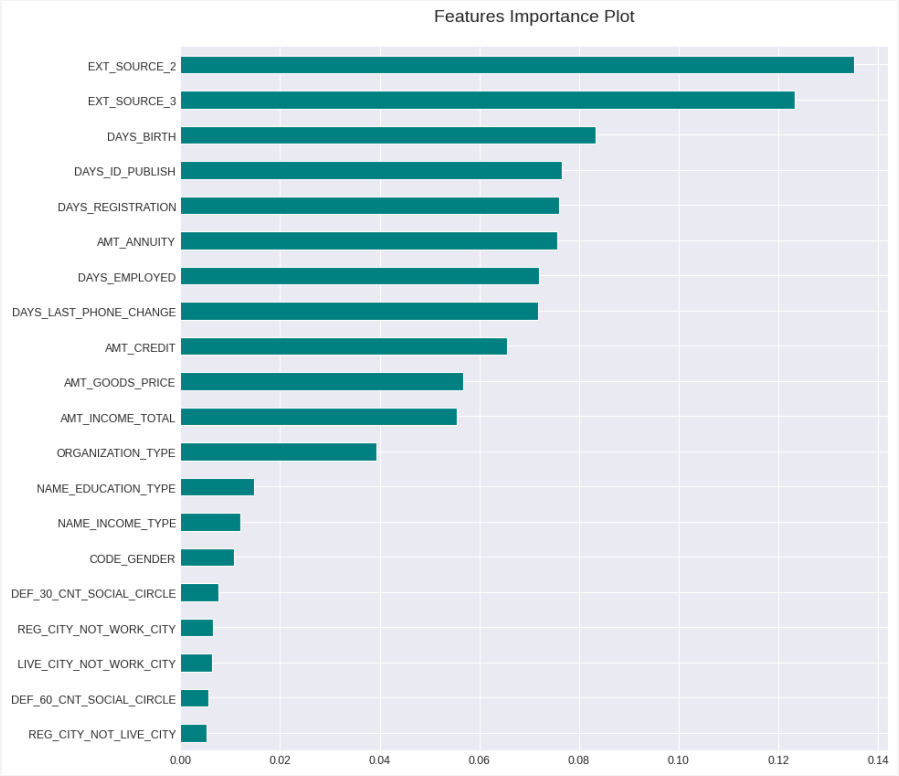
Without Hyperparameter Tuning

Algorithm	Training Accuracy Score	Testing Accuracy Score	ROC Score
Logistic Regression	67.16%	67.29%	67.28%
Random Forest	100%	99.65%	99.65%

With Hyperparameter Tuning

Algorithm	Training Accuracy Score	Testing Accuracy Score	ROC Score
Logistic Regression	67.16%	67.28%	67.28%
Random Forest	99.52%	95.33%	95.34%

FEATURE IMPORTANCE



Top 3 important features

- Score from external data source 2
- Score from external data source 3
- Client's age in days

MODEL SELECTION

By comparing 2 machine learning models between Logistic Regression and Random Forest, the best model that has the highest accuracy is Random Forest.

BUSINESS RECOMMENDATION

Prioritize married customers, as they tend to have high credit ratings and more stable financial situations, it may be beneficial to provide incentives or promotions to encourage them to utilize your company's financial services.

It's important to maintain good relationships with business people who have revolving loans and a positive payment history, as they can be valuable customers who may be interested in additional financial products or services.

Properly serve widow customers, who often have lower incomes and medium credit ratings with payment difficulties, it's crucial to evaluate their financial situations thoroughly and offer appropriate financial products and services.

THANK YOU!

https://github.com/ilutysn/VIX_HCI_Scorecard