HOME CREDIT SCORECARD MODEL

Michelle Natasha Vania Winarko

Background

- The difficulty of getting a loan due to sufficient credit history makes it difficult for many people to get a loan.

- Home Credit Indonedia wants to predict a client's credit repayment ability using a machine learning model.

- Using existing data, Home Credit Indonesia wants to know the criteria for customers who have no problems in the credit process that will help clients succeed in running their bussiness and also help increase their income.

Goal

- To identify clients who may have difficulty repaying loans and who may not.

- To predict client's repayment abilities

Dataset

- The data is provided by the Home Credit Indonesia team.

- It contains all customer data with problems and those who have no problems paying off loans.

Steps

1. Data Cleaning
2. Data Visualization
3. Build ML models
4. Predict client's repayment abilities
5. Provide recommendation for the company to increase their clients succeed in applying loans.

Data Preprocessing (application_train)

1. **Exploratory Data Analysis (EDA): discover patterns and the structure of the dataset.**
   - Bivariate Visualization (between 2 features)
   - Multivariate Visualization (more than 2 features)

2. **Data Cleaning**
   - Detecting Duplication
   - Handling Missing Values
   - Detecting Outliers

3. **Model Building**
   - Label Encoding: transforming non-numerical labels into numerical.
   - Feature Selection: identify the top 20 best features to include in the model.
   - Handling Imbalanced Data
   - Model Building
   - Model Evaluation: find the best model by comparing it.

Data Preprocessing (application_test)
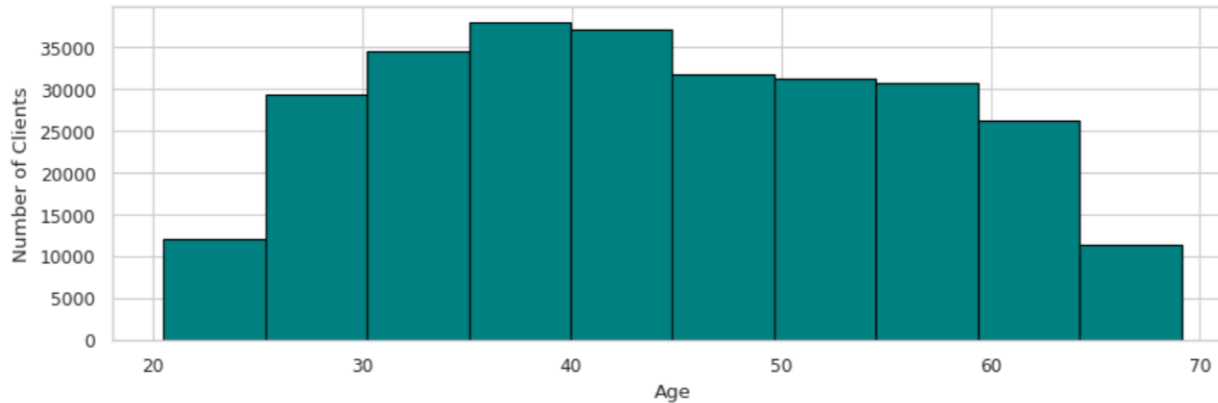
1. **Data Cleaning**
   - Detecting Duplication
   - Handling Missing Values
   - Label Encoding

2. **Prediction**
   - Predict client's repayment abilities with best machine learning model obtained before (with application_train)
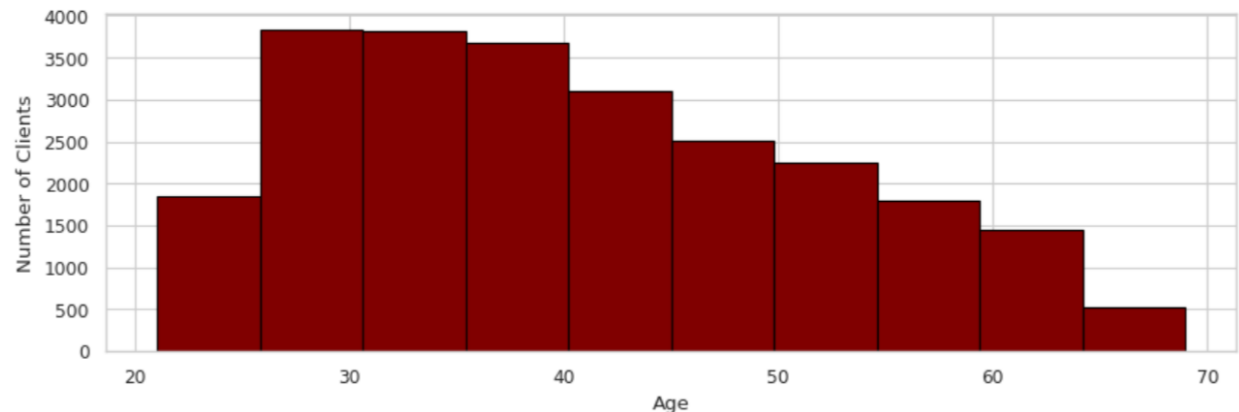
Business Insights

Age of Client (in years) who have No Payment Difficulties
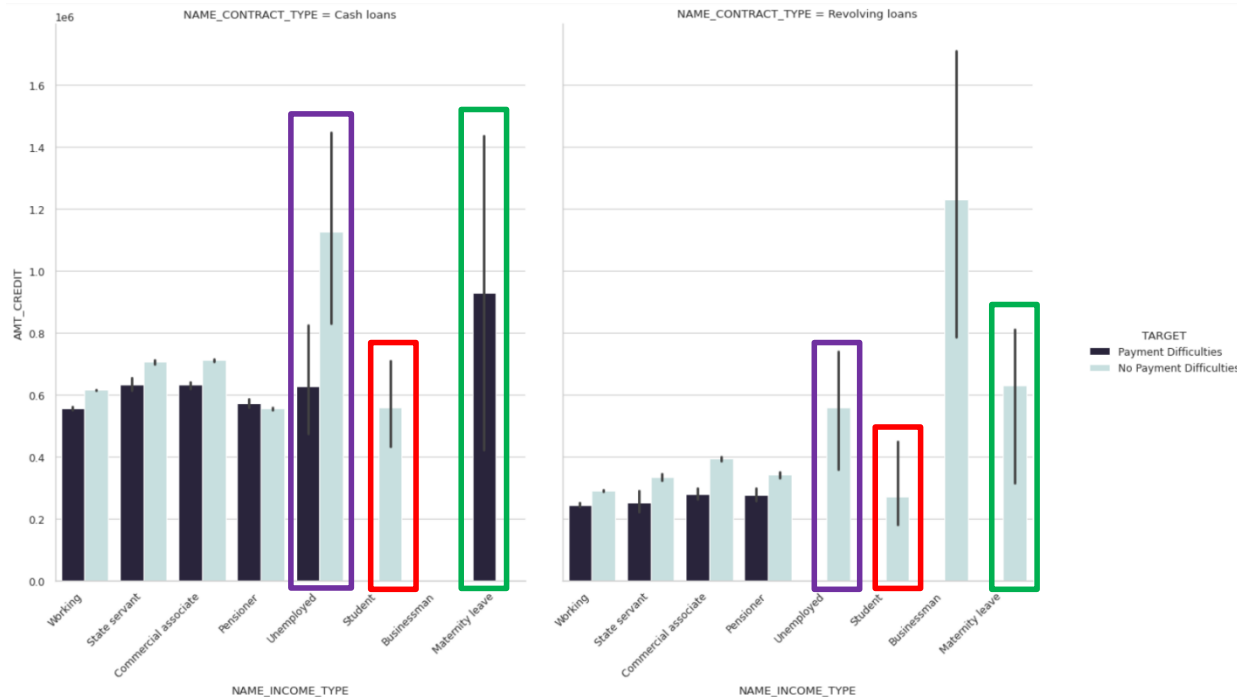

- Clients aged **35-40 years** are the highest age range who apply for loans.

- Clients aged **<25** or **>65** are the least age range who apply for loans.

- Clients aged **35-45 years** has no payment difficulties.

- Clients aged **25-35 years** have payment difficulties.

Age of Client (in years) who have Payment Difficulties

Business Insights

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

- For **maternity leave clients with cash loans**, clients **have problem repaying the loans fro a medium credit card amount of the loan**. While all maternity leave clients with revolving loans have No. difficulty repaying the loan.

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

Model Comparison

Algorithm	Training Accuracy Score	Testing Accuracy Score	Error Margin	ROC Score
Logistic Regression	67.16%	67.29%	0.13%	0.6729
Gaussian Naive Bayes	60.24%	60.39%	0.15%	0.604
Decision Tree	100.0%	88.33%	11.67%	0.8833
Random Forest	100.0%	99.65%	0.35%	0.9965
K-Nearest Neighbor	91.56%	88.07%	3.49%	0.8806
Neural Network	69.75%	69.12%	0.63%	0.6912

- **Random Forest** model has most accurate value than others where the training & testing accuracy score is not too different. It means there is **no underfitting or overfitting**.
- So, **Random Forest** model was chosen as the **best model to predict client's repayment abilities**.

Recommendation

• Make some promotional activities to attract more **students, accountant, high skill tech staff**, and **manager** to apply for a loan. Based on data, they are **capable of repaying the loans** but there are only a few applications that come from them.

• For **maternity leaves, cash loan**, and **unemployed** clients who is **incapable of repaying the loan** need to do a survei so that the company can recommend the right contract type in the future.

Project Documentation

- Github:

  https://github.com/michellenatashavw/Home-Credit-Scorecard-Model

THANK YOU!