HOME CREDIT SCORECARD MODEL

BY MUHAMMAD WILDAN VICKI
FAHMI AKBAR

01

PROBLEM RESEARCH

PROJECT BACKGROUND

- Many people struggle to get loans due to insufficient or non-existent credit histories. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience. Home Credit makes use of a variety of alternative data to predict their clients' repayment abilities. Doing so will ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

DATA
PROCESSING

02

Data Application Train	**122** Number of Columns	**307,511** Number of Rows

EDA
Discover patterns, and the structure of the dataset

→ **DATA CLEANING**

→ **MODEL BUILDING**

Raw Data Application Train

Bivariate Visualization

Visualization of the relationship between 2 features

Multivariate Visualization

Visualization of the relationship of more than 2 features

Detecting Duplication

No duplicate rows

Handling Missing Values

There are some columns that are dropped and the rest are imputed

Detecting Outliers

There are some columns that have outliers, but it was decided the outlier will not be removed

Label Encoding

Transform non-numerical to numerical labels

Feature Selection

Identify the top 20 best features to include in the model

Handling Imbalanced Data

Re-sampling so that the data is balanced

Model Building

Build models with multiple machine learning algorithms and compare which one is the best
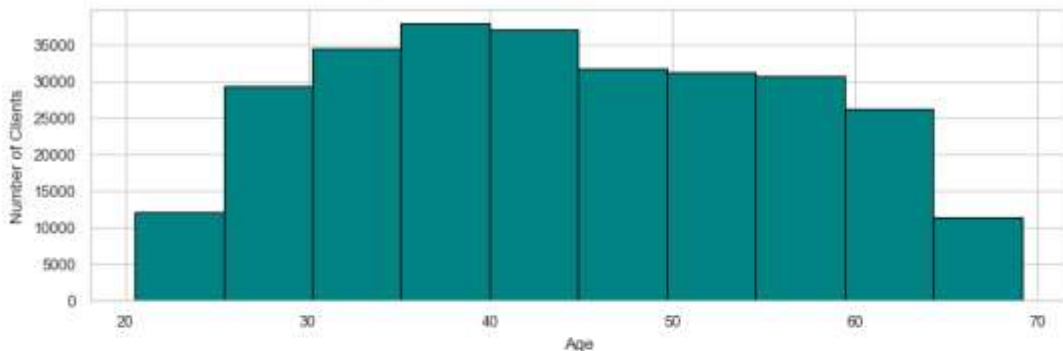
Model Evaluation
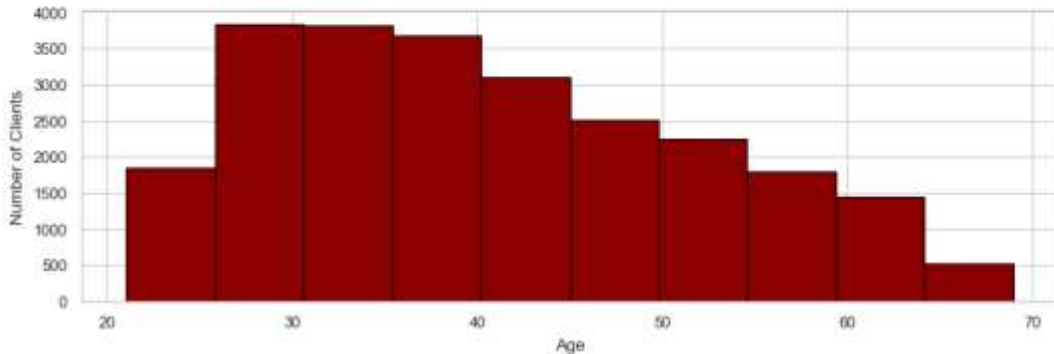
Compare which one of the model is the best

BUSINESS INSIGHT

03

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

Age of Client (in years) who have Payment Difficulties

- Most number of clients who apply for loans are in the range of 35-40 years.
- Meanwhile, the number of applicants for clients aged <25 or age >65 is very low.

- Clients who have **no payment difficulties** are clients in the range of 35-45 years. You can target these clients as your priority.

- While clients who **have payment difficulties** are client the range of 25-35 years.

04 MACHINE LEARNING MODEL

Algorithm	Training Accuracy Score	Testing Accuracy Score	Error Margin	ROC Score
Logistic Regression	67.16%	67.29%	0.13%	0.6728
Gaussian Naive Bayes	60.24%	60.39%	0.15%	0.604
Decision Tree	100%	83.9%	11.74%	0.8826
Random Forest	100%	99.63%	0.37%	0.9963
Neural Network	69.59%	68.94%	0.65%	0.6894

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, which is there is **no underfitting or overfitting**. So the **Random Forest** model was chosen as the **best model** to **predict client's repayment abilities**.

RECOMMENDATION

1. A client with an income type of **student** can be said to be a client who is **capable of repaying the loans** whether with a cash loan or revolving loan (100% of applications approved). But there only 0.005% of applications come from the student.

2. A client who works as an **accountant** can be said to be a client who is **capable of repaying the loans** (95% of applications approved). But, there is only 3.19% of applications come from an accountant. So do, the client who work as **high skill tech staff** and **manager**, they are capable of repaying the loans, but there are only a few applications

THANK YOU