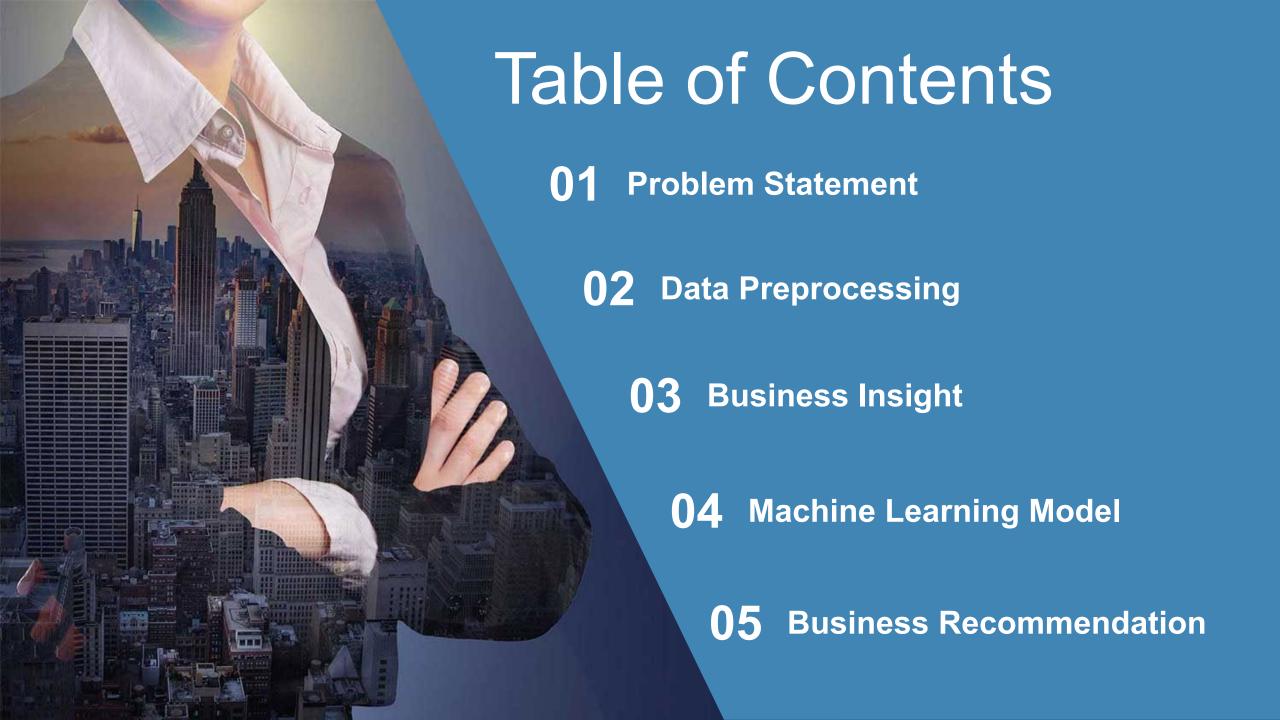


Home Credit Scorecard Model

by Fikrie Lazuardi Imaduddin





Project Bacground

People struggle to access credit, which is not due to a lack of trying. Even individuals who earn more than \$100,000 annually may live paycheck to paycheck, making it difficult for them to access credit when they need it. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. Home Credit are trying to make sure this underserved population has a positive loan experience by applying credit scoring card.



Data Source

The data used are application train and application test. There are our main table, broken into two files for train (with TARGET).



Objective

- 1. Identify characteristics of potential clients who will have difficulty repaying loans.
- 2. Predict client's repayment abilities.

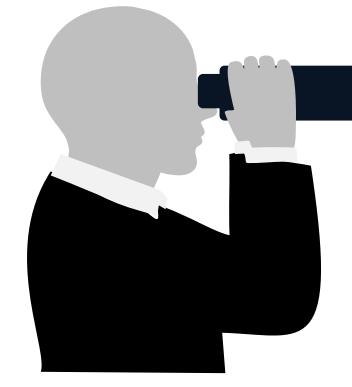


Actions

- Perform data cleaning, and visualization for business insights
- 2. Build a models with machine learning algorithms



- 3. Predict client's repayment abilities on application data test with best machine learning model
- 4. Provide recommendations for company to increase their clients succeed in applying for loans



Data Preprocessing



EDA

Discover patterns, and the structure of the dataset.



Data Cleaning

Process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.



Model Building

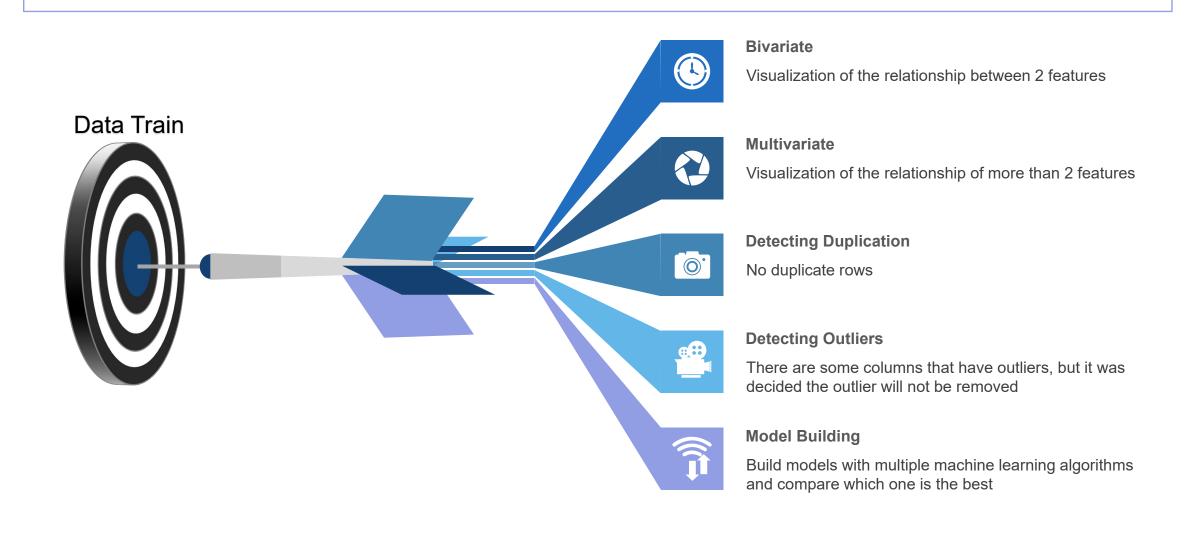
Machine learning takes a static system and its data and makes it dynamic by learning from previously collected information from sensors, measuring devices and occupant behaviors.

PREDICTION

Predict client's repayment abilities with best machine learning model obtained before

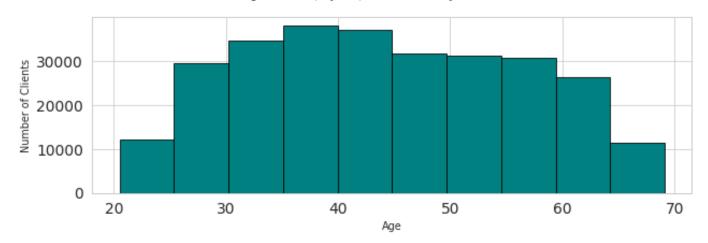
Data Application Train

122 Number of Columns 307,551 Number of Rows

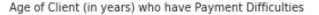


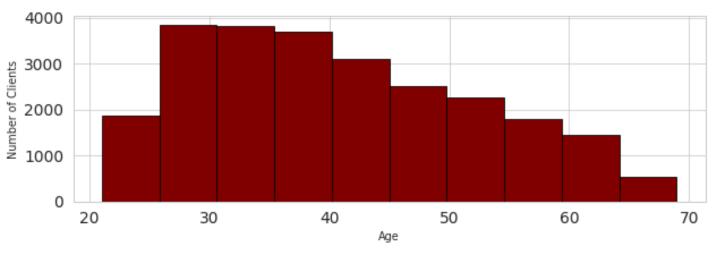


Age of Client (in years) who have No Payment Difficulties



- 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.



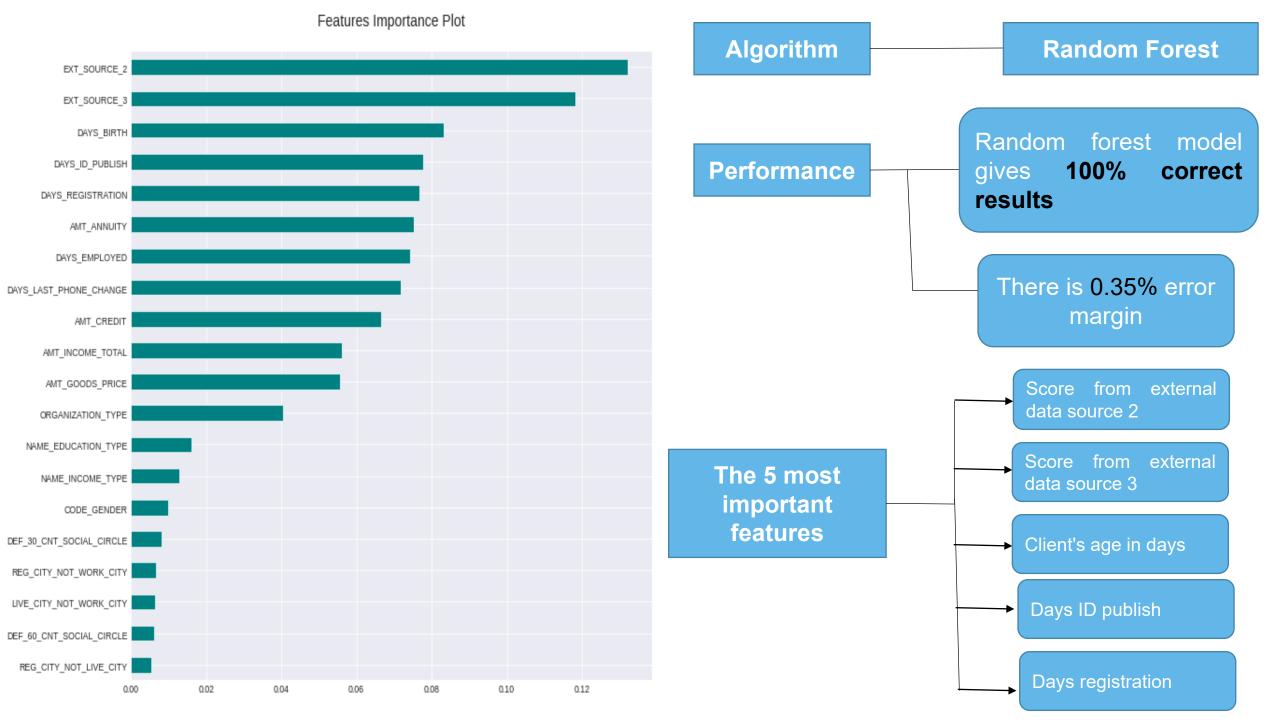


- Clients who have no payment difficulties are clients in the range of 35-45 years. You can target these clients as your priority.
- While clients who have payment difficulties are client the range of 25-35 years.



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%	99.65%	0.35%	0.9965
K-Nearest Neighbor	91.56%	88.07%	3.79%	0.8806

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.



Business Recommendation



Recommendation

Adding more campaign so people such as student, a ccountant, hi gh skill tech s taff, manager interested in applying for a loan.

It could be a survey to find out if there is a problem if a client with maternity lea ves or unem ployed takes a cash loans contract.

