

Home Credit

Indonesia Score

Card Model

Nicholas Sky Salvatio

Problem Statement

Based on a some time period loan applications dataset, Home Credit Indonesia discover that **there are numerous customers who were struggling to get loans due to insufficient or non-existent credit histories.** Therefore, Home Credit Indonesia strives to broaden financial inclusion for the unbanked population by providing a safe and positive borrowing experience. Home Credit Indonesia would like to develop machine learning model to help the company determine whether the proposed loan is accepted or rejected by identifying the possible significant factors. Using the model, Home Credit Indonesia could **predict their customers' payment abilities and ensure that customers capable of payment will not be rejected** and that the loans are given with a principal, maturity, and repayment calendar that will empower them to be successful.

Objectives:

- Identify the significant factors which cause the risk of loan payment difficulties.
- Implement machine learning algorithm to predict customer's loan payment abilities.
- Provide actionable solutions to prevent and reduce the number of customers who have loan payment difficulties.

Actions:

- Perform Exploratory Data Analysis (EDA) to gain data and business insights.
- Perform Data Preprocessing before continuing to the model building.
- Develop machine learning models to predict customer's payment abilities.
- Determine the best machine learning model with the highest prediction metrics score.
- Provide recommendations for the company to reduce the failures of the customers in applying loans.

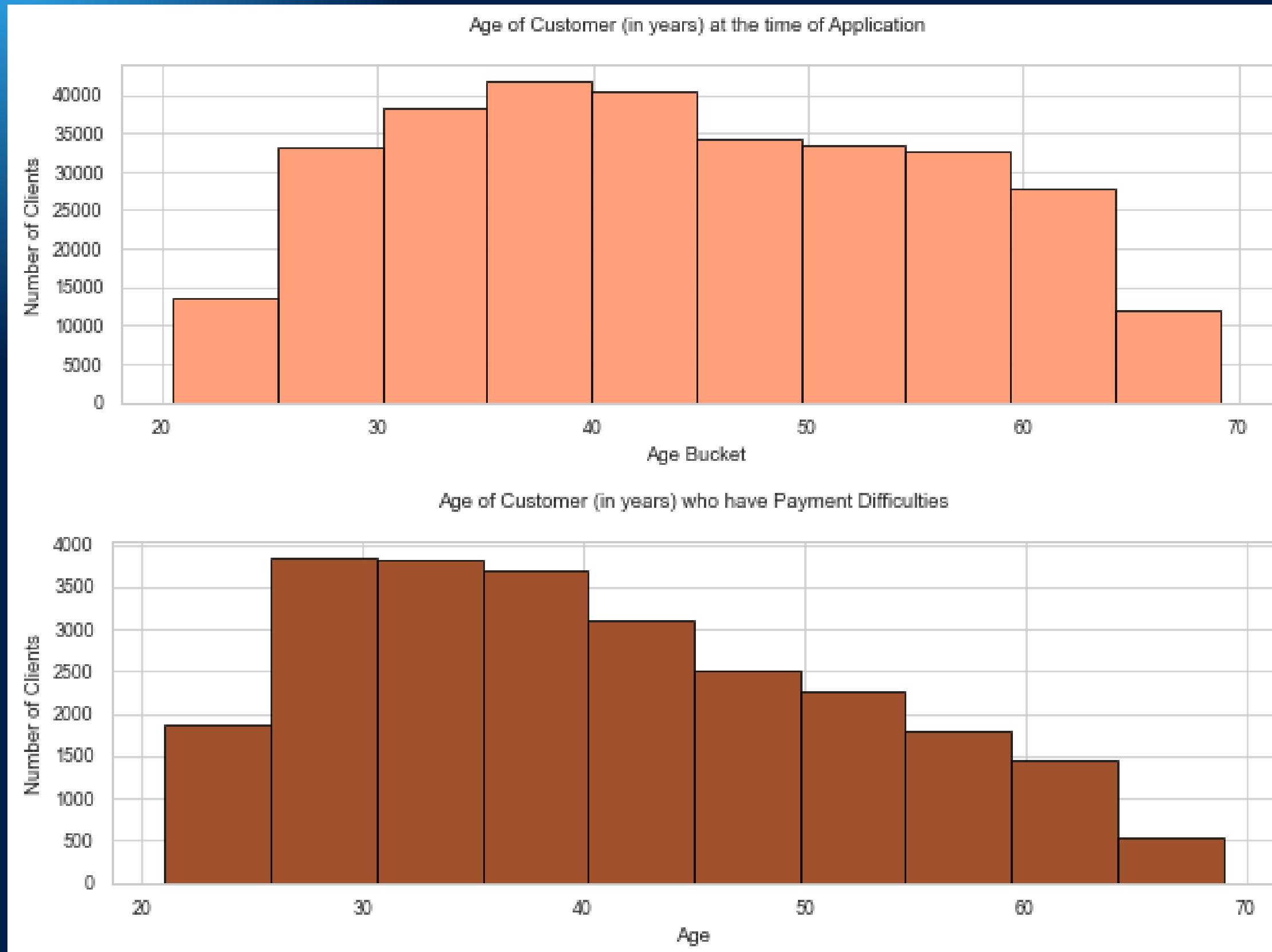
Dataset

The data that will be used for this project are **Application Train** and **Application Test**. These two datasets are our main table, broken into two files; one for **training (with TARGET)** and the other for **testing (no TARGET)**.

Data Preprocessing

- 1.Exploratory Data Analysis (Univariate, Bivariate, and Multivariate Analysis)
- 2.Handling duplication, missing values, and outliers.
- 3.Label Encoding and Feature Selection/Feature Engineering
- 4.Handling Imbalanced Data by Resampling
- 5.Model Building and Model Evaluation

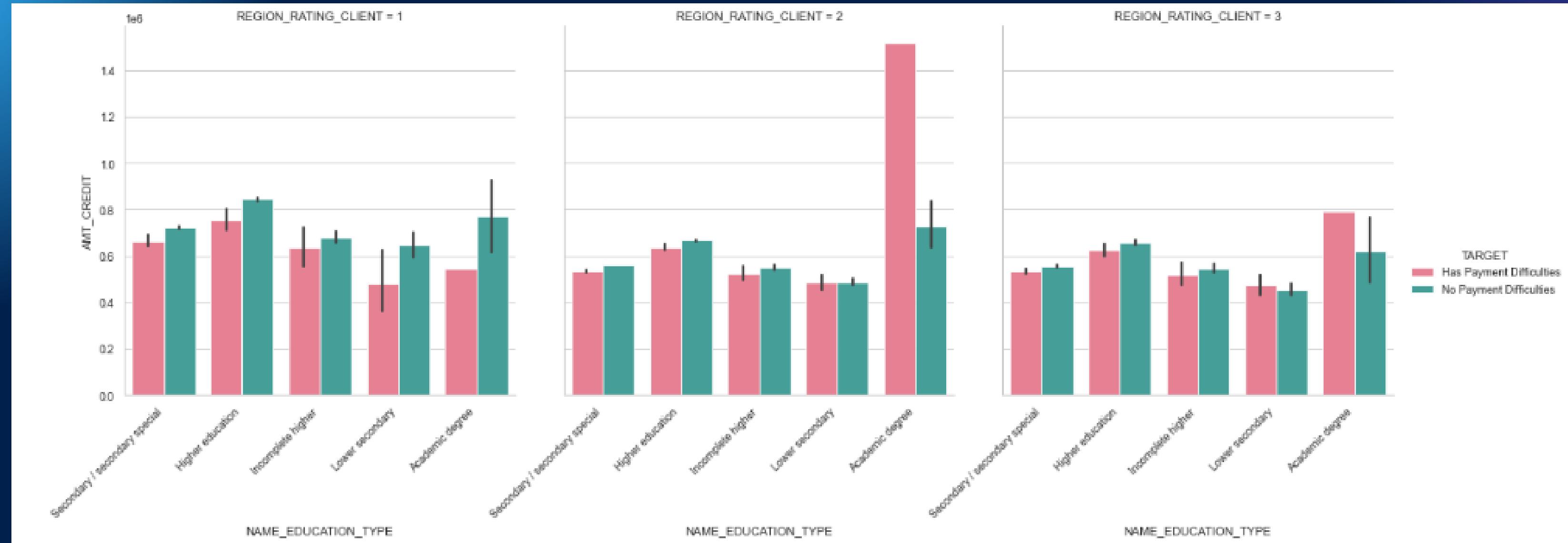
Business Insights



The majority of the customers who applied for loans are in the range of 30-45 years old, while the minority are the customers aged less than 25 and greater than 65.

Customers who have payment difficulties are ranged between 25-35 years old. Most of them have income type of Maternity Leave.

Business Insights



Customers who have highest education of Academic degree and are living in a **region with rating of 2 and 3** tend to **have more payment difficulties** than the others for **moderate to high amounts credit of the loans**.

Machine Learning Model Comparison

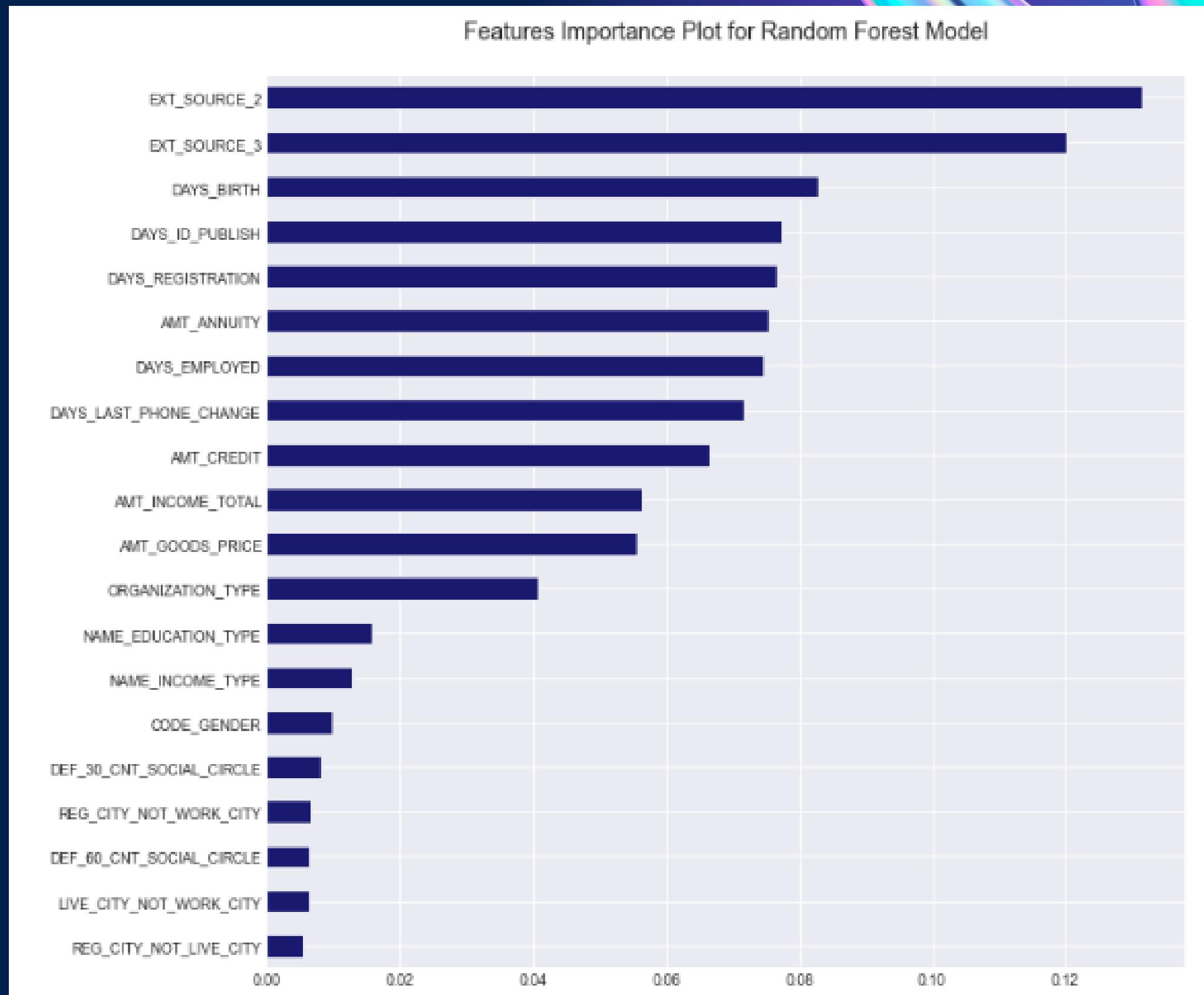
No.	Model	Accuracy	Precision	Recall	F1-Score	AUC
1	Random Forest	0.9962	0.9930	0.9994	0.9962	0.9962
2	Decision Tree	0.8835	0.9014	0.8614	0.8810	0.8835
3	K-Nearest Neighbor	0.8807	0.8093	0.9962	0.8931	0.8806
4	XGBoost Classifier	0.7183	0.7262	0.7016	0.7137	0.7183
5	AdaBoost Classifier	0.6767	0.6881	0.6472	0.6670	0.6767
6	Logistic Regression	0.6729	0.6772	0.6614	0.6692	0.6729
7	Nearest Centroid	0.6516	0.6513	0.6533	0.6523	0.6516

Best Model: Random Forest

The best model that is selected for prediction is the **Random Forest** model with the relatively **highest metrics score** (Accuracy, Precision, Recall, F1-Score, and AUC) among other models.

The **top 5 most important features** from the **Random Forest** model are:

1. Score from external data source 2.
2. Score from external data source 3.
3. Customer's age in days
4. Days ID published
5. Days registration



Business Recommendations

1. It is recommended to conduct **further analysis** (ex: surveys) to discover if there are problems/difficulties for customers with young age (25-35 years old), maternity leaves, has Academic degree, and currently living in specific regions from taking loans contract. Therefore, in the future, we could **recommend the right contract type** for them so that their applications will be approved.
2. It is recommended to establish **Customer Segmentation Analysis** to classify customers based on their transactional activities/behaviors in order to **discover the hidden patterns of the characteristics of the customers** who are having/not having payment difficulties so that we could suggest actionable strategies and solutions to enhance business metrics and performances.

You can view the entire
documentations here:

<https://github.com/nicholassky0103/Home-Credit-Indonesia-Score-Card-Model>

Thank You