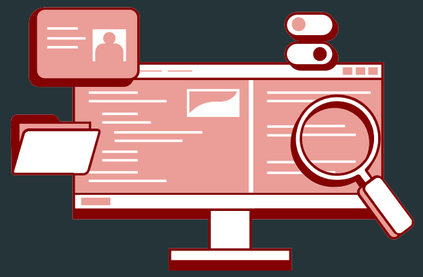
HOME CREDIT

HOME CREDIT
SCORECARD MODEL

by Vania Elvina Afiati

TABLE OF **CONTENTS**



PROBLEM RESEARCH

PROJECT BACKGROUND

Home Credit is an international consumer finance provider focusing on responsible **lending**. Currently, Home Credit is utilizing various statistical and machine learning methods to **predict credit scores.** In this project, we want to ensure that customers capable of repaying loans are not rejected when applying for one, and loans can be provided with a principal, maturity, and repayment calendar that will motivate customers to succeed.

Objective 1

Identify traits of potential clients who are likely or unlikely to face payment difficulties.

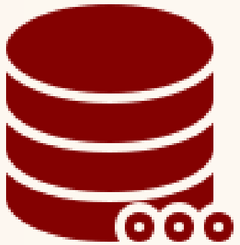
Objective 2

Forecast clients' ability to meet repayment obligations.

Objective 3

Suggest business recommendation based on the findings.

DATA PRE-PROCESSING



Raw Dataset

Exploratory Data Analysis (EDA)

Gathering information and visualizing the dataset.

It was obtained that the dataset has some imbalances and need to be resampled.

Data Cleaning

Checking missing values and duplicates in the dataset.

It was obtained that there were no duplicate values. However, some missing values were found and handled by removing columns with more than 50% missing values and replacing some others with median values.

Feature Selection & Engineering

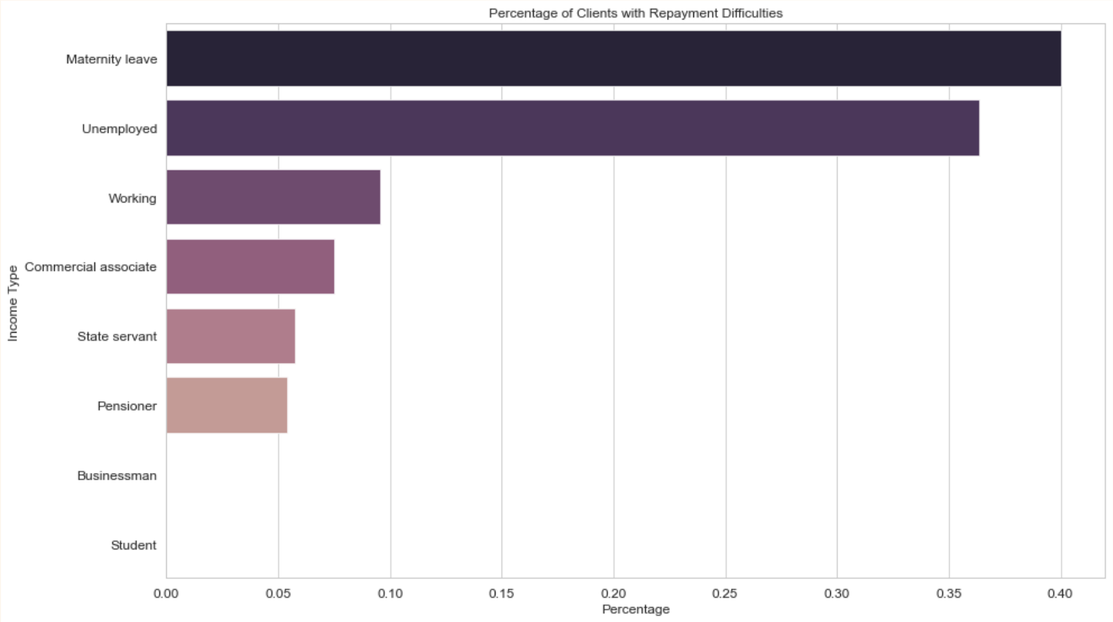
Analyzed features correlation and select features with highest correlation. It was obtained that clients'employed days have the highest correlation with ability to repay loan.

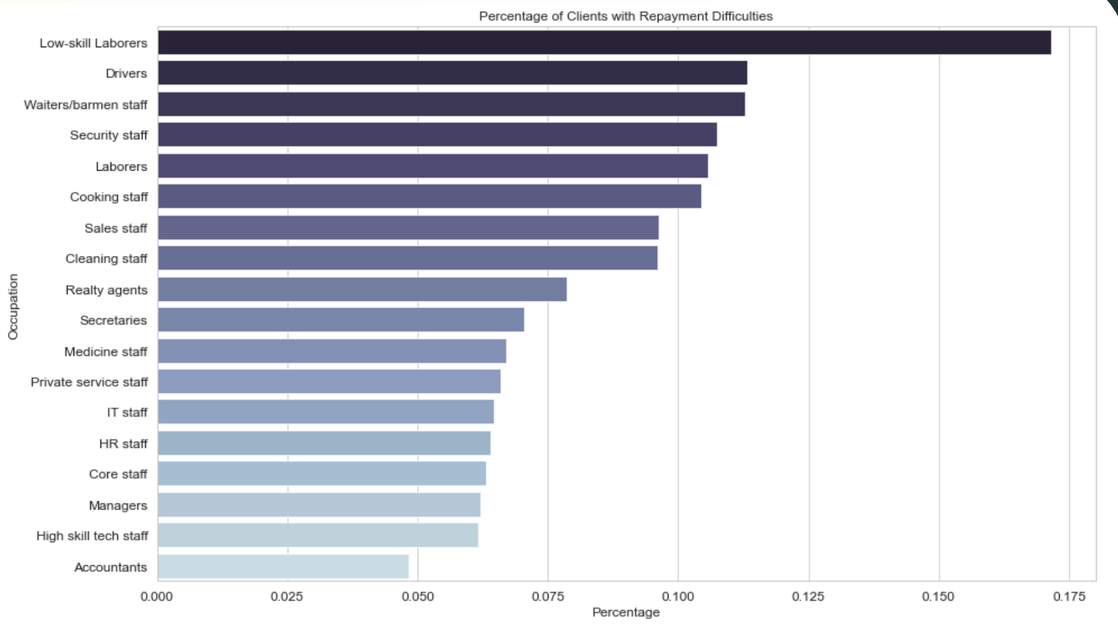
Split Dataset

Splitting dataset for training and testing.

Model

DATA VISUALIZATION





From the graph above, it is obvious that **Student** and **Businessman** have no difficulties in repaying their loans. On the other hand, clients who are in **Maternity Leave** are the highest percentage of clients with difficulty repaying the loans. Additionally, among the employed clients, **Low-skill Laborers** are the highest percentage of clients with difficulties repaying the loans, while **Accountants** are the lowest.

MACHINE LEARNING MODEL

Model	Training Accuracy	Testing Accuracy
Logistic Regression	0.6716	0.6729
Random Forest	1	0.9965
K-Nearest Neighbors	0.9156	0.8807

Based on the model evaluation above, it is obvious that the best machine learning model to analyze Home Credit client data is **Random Forest.**

BUSINESS **RECOMMENDATION**

1 TARGET KEY DEMOGRAPHICS

Launch tailored campaigns to attract **students** and **businessman** by showcasing features of loan products that aligned with their needs.

2 OFFER ALTERNATIVES

Recommend clients on **maternity leave** and **unemployed** to use suitable contract types to improve approval rates.

Thank You

Codes for this project can be accessed in:

https://github.com/vaniaelvina/hci-credit-scorecard-model