**Wika Rabila Putri**

rabilaputri@gmail.com

https://www.linkedin.com/in/wika-rabila-putri/

Bandung, Indonesia

Fresh graduate with a keen interest in data and finance. Armed with hands-on experience in SQL, Python, Excel, and data visualization from rigorous projects and a specialized bootcamp, Wika excels at turning complex data into actionable insights. With a solid foundation in financial management and a strong drive for data-driven decision-making, Wika is eager to contribute to a forward-thinking team as a Business Analyst. Her goal is to leverage her analytical skills and passion for data to help optimize business strategies and drive impactful results.

# Project **Portfolio**

A financial services firm seeks to improve its credit risk evaluation and management strategies. The objective is to enhance loan approval processes and minimize financial losses due to defaults. To accomplish this, we are building an advanced machine learning model that utilizes data from both granted and declined loans.

This model will examine historical loan data to more accurately predict credit risk. By integrating a diverse dataset that encompasses detailed insights on both approved and denied loan applications, the model aims to uncover patterns and indicators associated with high-risk borrowers.

Project explanation video: **Click here!**

Code: **Click here! (GitHub)**

# About **Company**

Home Credit Indonesia, established in 2013, is a leading consumer finance company providing fast and easy financing solutions. Headquartered in Jakarta, the company offers a wide range of services including consumer loans, cash loans, and credit cards. Their mobile app allows customers to access financing quickly and securely. With a vision to improve the quality of life for Indonesians through responsible lending, Home Credit Indonesia has reached over 10 million customers, working with thousands of retail partners across the country. The company is committed to innovation, financial literacy education, and secure data protection to enhance customer experience and drive financial inclusion.

## Rakamin Academy

# Table of **Content**

# Business Understanding

# Current Problem.

Home Credit faces challenges in accurately assessing credit risk, leading to the rejection of customers who are financially capable of repaying loans. Current models may not fully capture all relevant factors, resulting in missed opportunities for lending and potential customer dissatisfaction.

## Impact:

Home Credit's challenges in accurately assessing credit risk lead to the rejection of financially capable customers, resulting in lost revenue, customer dissatisfaction, and a damaged reputation. This creates opportunities for competitors to gain market share and can lead to inefficient resource allocation. Additionally, inaccurate assessments may increase default risks, regulatory issues, and operational costs. Improving credit risk assessment models is crucial for maintaining profitability and customer satisfaction.

# Goals.



Develop a more precise credit scoring model using Machine Learning techniques, create a reliable scorecard that improves loan performance, enhances customer satisfaction, and supports responsible lending practices.

# Data Overview.

1. The dataset provided consists of **307.511 entries** and **122 columns**

2. The dataset consists of **106 numeric** columns and **16 categorical columns**

3. There are **67 columns** with missing (null) data in the dataset

4. No duplicate data

- Features with **more than 40%** missing data are removed to avoid unreliable imputation, reduce bias, and improve model performance by maintaining data quality and simplicity.
- To handle anomalous values we replaced the value with the mode to maintain data consistency.
- Outliers are removed by calculating the Z-scores of the data and filtering out values that are more than 3 standard deviations away from the mean.

## Percentage of Customers Who Apply For Loans

## Distribution of Loan Repayment Status by Marital Status
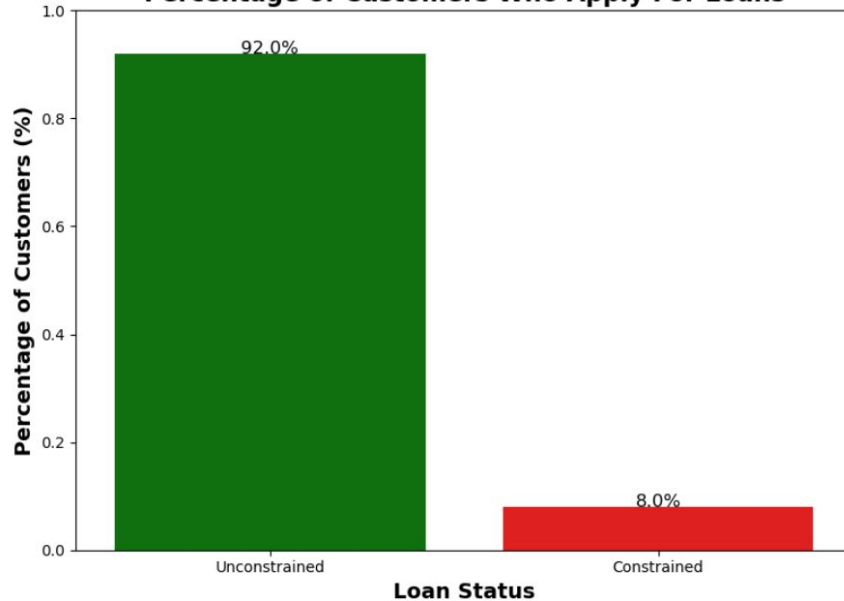
**92%** of customers managed to repay their loans **on time**, while **8%** encountered difficulties. This notable discrepancy leads us to investigate the factors that set these two groups apart, specifically the characteristics that contribute to timely loan repayment.
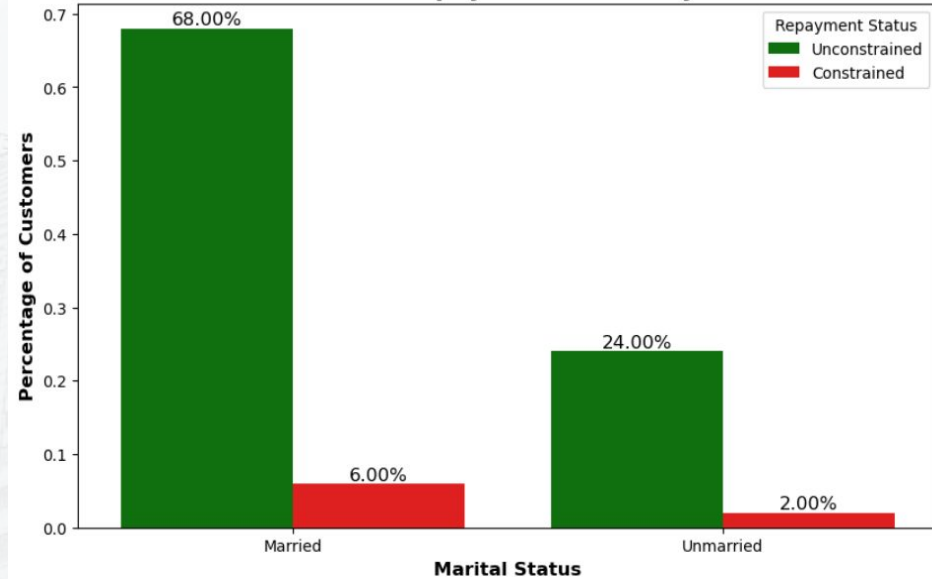
This suggests that being married may be associated with more reliable loan repayment behavior.

**Top 6 Occupations with Highest Unconstrained Customers Percentage**
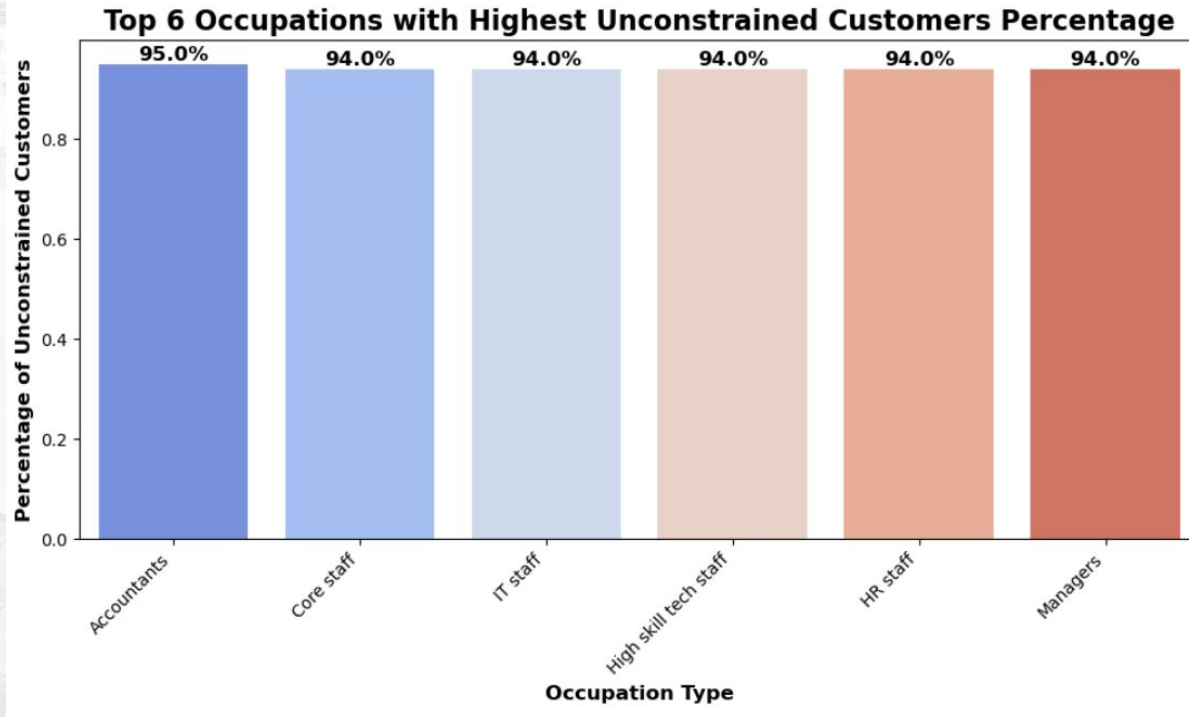
- Occupations such as Accountants, Core Staff, IT Staff, High-skill tech staff, HR Staff, and Managers show a high percentage of Unconstrained customers (94% or above), meaning most individuals in these roles do not face issues repaying loans.

# Data Preparation.

**Feature Creation**

**Drop Column**
With more than 40% missing values, Unnecessary columns

**Impute Column**
With less than 40% missing values, Numerical: median, Categorical: mode

**Feature Selection**
Selects columns with >two unique values & corrects negative values to absolute values

**Feature Engineering**
Encoding, Scaling, SMOTE

| Model | Data Split | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|---|
| Logistic Regression | Train (Oversampling) | 0.94 | 0.95 | 0.94 | 0.94 |
| | Test (Oversampling) | 0.94 | 0.95 | 0.94 | 0.94 |
| Logistic Regression | Train (Undersampling) | 0.68 | 0.68 | 0.68 | 0.68 |
| | Test (Undersampling) | 0.68 | 0.68 | 0.68 | 0.68 |
| XGBoost | Train (Oversampling) | 0.96 | 0.96 | 0.96 | 0.96 |
| | Test (Oversampling) | 0.95 | 0.96 | 0.95 | 0.95 |
| Random Forest | Train (Oversampling) | 1.00 | 1.00 | 1.00 | 1.00 |
| | Test (Oversampling) | 0.96 | 0.96 | 0.96 | 0.96 |

**Random Forest (Oversampling)** as the best model because it has the highest testing accuracy of 0.96 and a good balance in precision, recall, and f1-score, all at 0.96.

# Summary

# Summary.

- The Home Credit Scorecard Model analysis identifies key customer segments for revolving loan contracts, focusing on individuals with income types such as business owners, maternity leave recipients, students, and unemployed individuals to tailor effective loan offerings.
- Recommendations include launching campaigns targeting accountants, the largest customer group with the best loan repayment rates, as well as creating targeted advertisements for HR professionals, IT specialists, and real estate agents to boost loan applications.
- Following a thorough evaluation, the **Random Forest** machine learning model is selected as the primary tool for predicting customer loan payment issues. This model demonstrates robust performance in identifying potential risks, enabling proactive management of possible loan defaults.

# Thank You