

# Home Credit Scorecard Model

Project Based Internship Rakamin x Home Credit Indonesia



Kevin Haposan Arip ratama

[Github](#)



# Contents



**01**

Problem Research

**02**

Data Preprocessing

**03**

Data Visualization and  
Business Insight

**04**

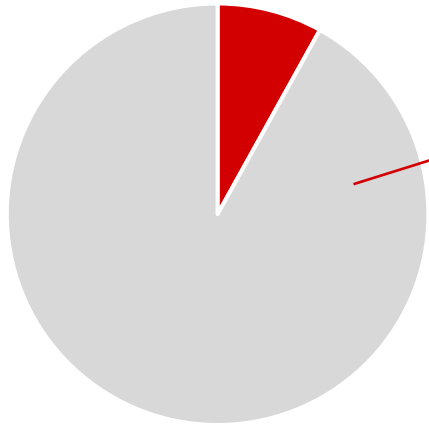
Machine Learning  
Implementation and  
Evaluation

**05**

Business Recommendation

# 01 Problem Research

Target Distribution (%)



**91, 93% of  
customers  
loans are repaid  
on time**

**Not Repaying  
Loans  
=  
Reducing the  
Company's  
Revenue**

Background

Home Credit Indonesia aims to determine customers who are capable or incapable of loan repayment when applying for a loan.

Goals

Reducing the rate of customers who are unable to apply for loans.

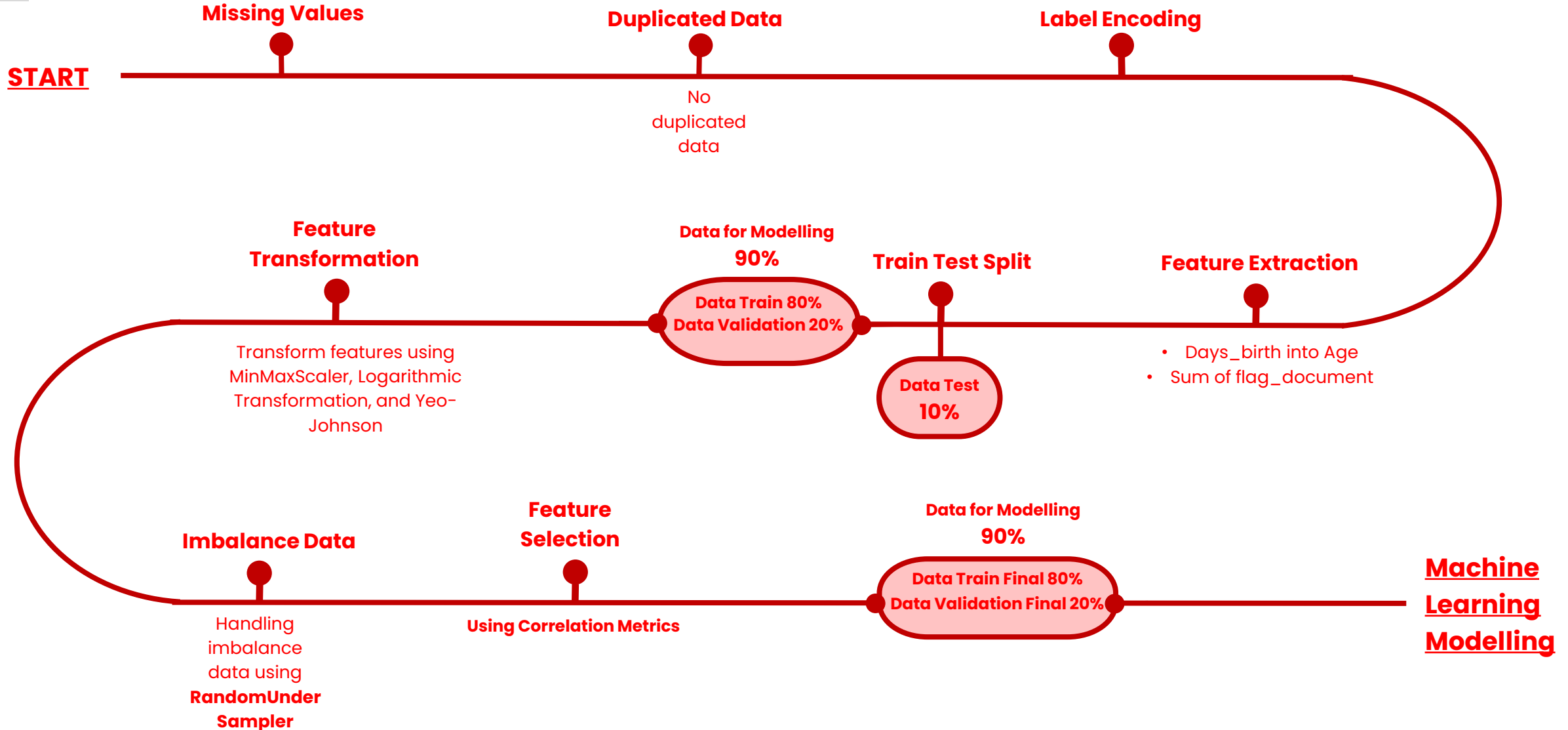
Objectives

Developing a machine learning model that can predict customers who are unable to repay their loans and the factors influencing it.

Business Metrics

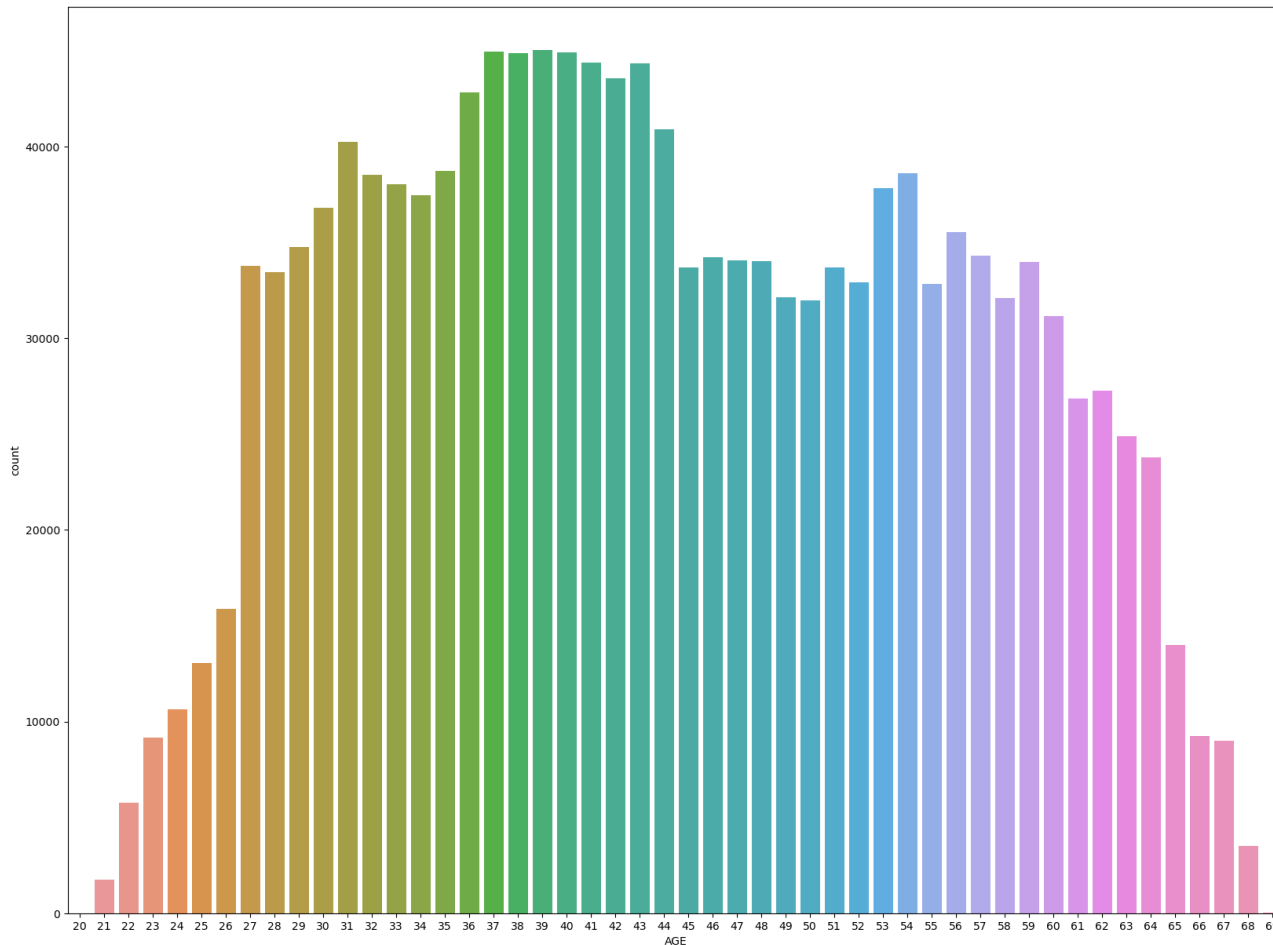
The percentage of customers unable to repay their loans (%).

## 02 Data Preprocessing



## 03 Data Visualization and Business Insight

The Distribution of Age of Customers Taking Loans



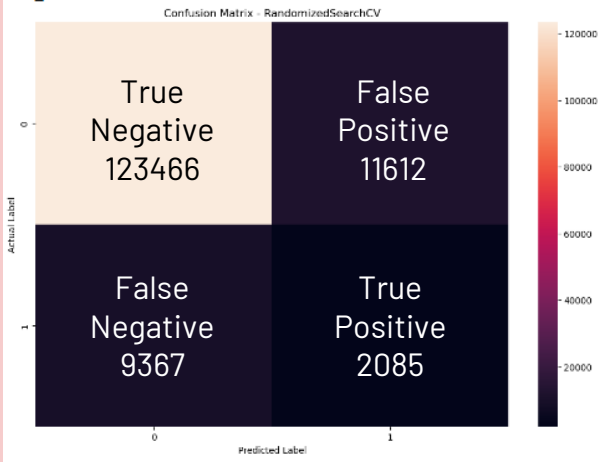
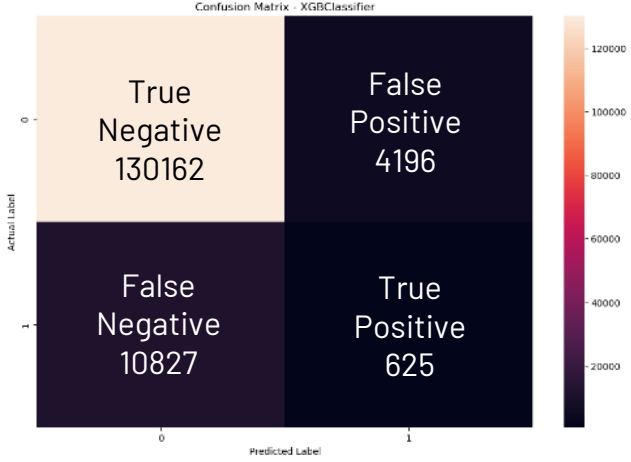
Customers who apply for a loan are generally between 27-64 years old, which is the typical age range for working individuals who have established financial responsibilities, such as supporting a family. This age group is often considered the most creditworthy, as they have steady income and a proven track record of managing their finances. By targeting this demographic, lending institutions can increase their chances of approving loans to creditworthy customers, reducing the risk of default. Additionally, offering loan products tailored to the needs of this demographic, such as family-friendly repayment terms or higher loan amounts, can help lenders differentiate themselves from competitors and build customer loyalty.

## 04 Machine Learning Implementation and Evaluation

MODEL	TUNING HYPERPARAMETER	ACCURACY (TRAIN)	ACCURACY (VALID)	RECALL (TRAIN)	RECALL (TEST)	AUC (TRAIN)	AUC (TEST)
Logistic Regression	Before	0,61	0,61	0,60	0,61	0,65	0,65
	After	0,61	0,61	0,60	0,61	0,65	0,65
Random Forest	Before	1	0,83	1	0,88	1	0,91
	After	0,67	0,65	0,67	0,65	0,74	0,70
XGBoost	Before	0,71	0,67	0,71	0,69	0,78	0,74
	After	0,60	0,58	0,61	0,62	0,63	0,63

After tuning hyperparameters, the Random Forest model and the XGBoost model, even before hyperparameter tuning, exhibit sufficiently high parameter values. Therefore, both models will be tested on the test set."

## 04 Machine Learning Implementation and Evaluation

Model	Random Forest	XGBoost
Confusion Metrics		
Accuracy	85,68%	89.26%
Recall	18,21%	5.46%
ROC_AUC	62,53%	54.62%

The Random Forest model was chosen because it has higher parameter values compared to XGBoost.

True Positive = 2085  
False Negative = 9367

Assumption 50%  
% Customer unable to apply for the loans =  $(0,5(TP) + FN) / \text{Total}$

Before Model = 8,07%  
After Model = 7,81%



## 05 Business Recommendation



Based on the identified business insight, here are some potential recommendations for lending institutions:

- Tailor loan products to the needs of working individuals aged 27-64, such as offering flexible repayment terms that align with their income patterns and family expenses.
- Consider offering higher loan amounts to this demographic, as they are more likely to have established financial responsibilities and a proven track record of managing their finances.
- Implement targeted marketing campaigns to reach this demographic, highlighting the benefits of your loan products and how they can help them achieve their financial goals.



—

# Thank You

