

# SCORECARD MODEL

BY : IMAM MAGHFIR RAMADHAN

VIX - VIRTUAL INTERNSHIP PROGRAM



<https://www.linkedin.com/in/imam-maghfir-ramadhan-5063b8183/>



<https://github.com/imammaghfir/ScoreCardModel/tree/main>

# 1.) PROBLEM RESEARCH



## PROBLEM STATEMENT

**Home Credit** aims to optimize its lending process by **minimizing the number of customers approved for loans who eventually default** on their payments. The company seeks to develop strategies or predictive models that effectively evaluate customer creditworthiness.



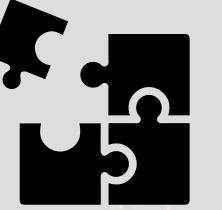
## GOAL

Minimize number of **customers** authorized for approval but who **ultimately default**



## OBJECTIVES

- **Identifying** the characteristics of **customer defaulters**
- **Building** a machine learning (ML) model that can **predict** customers who have the potential to become **customer defaulters**



## MODEL EVALUATION

### Precision

Accuracy of the model in predicting customer defaulters on a loan or making a purchase.

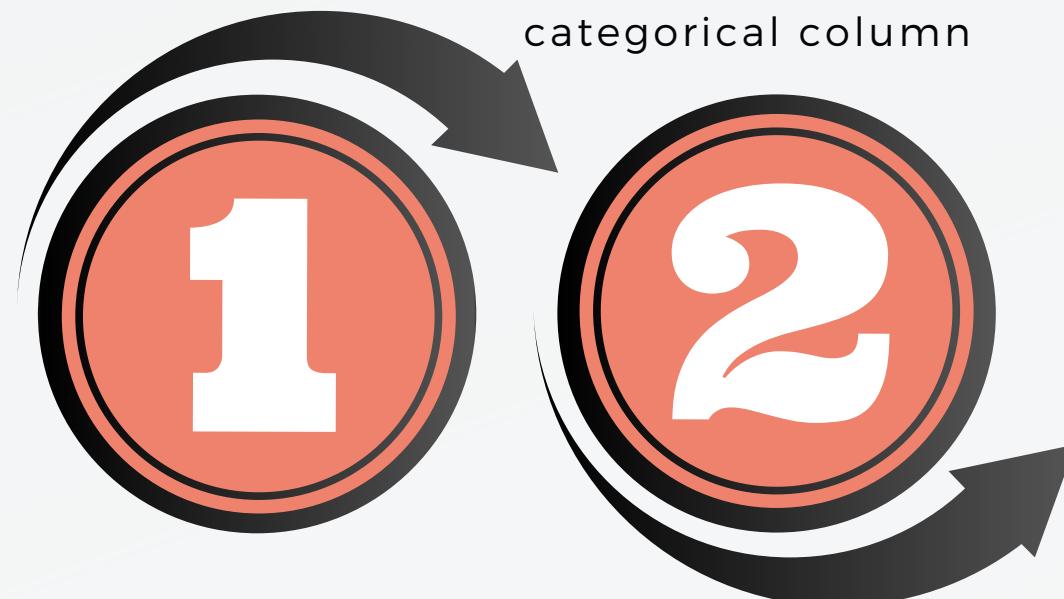
## 2.) DATA PRE-PROCESSING



### Handling Missing Value

There are 67 columns that contains **missing value**.

- **Dropping column** that has **>50%** missing value
- **Filling missing value** with **median** for numerical column and **mode** for categorical column



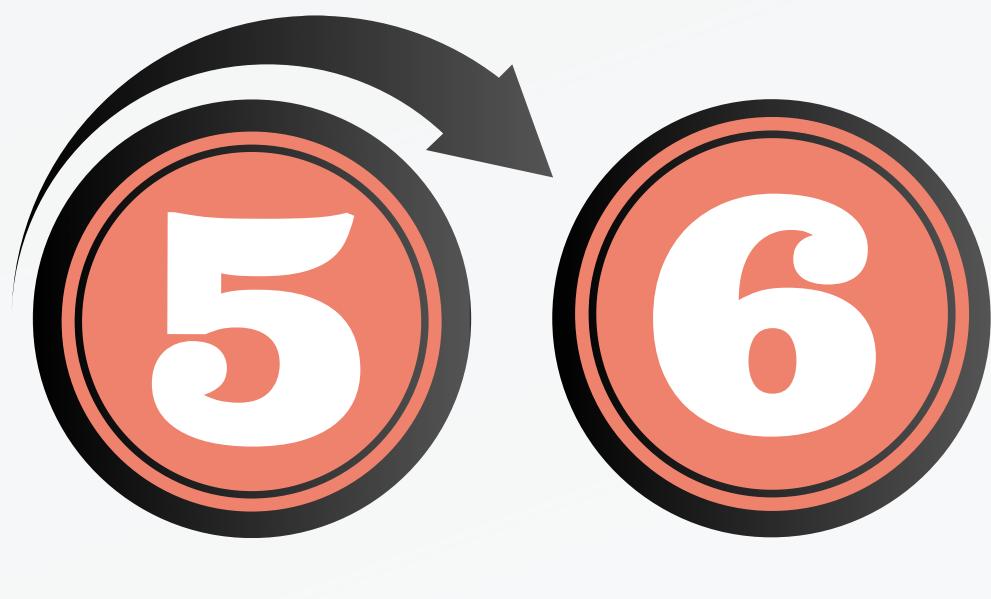
### Checking Duplicate Value

Based on the dataset, there is **no duplicate value found**

### Selecting Feature

All **features** will be used in the model (**except any ID features and redundant features**)

*Better result compare to use certain feature only based on experiment*



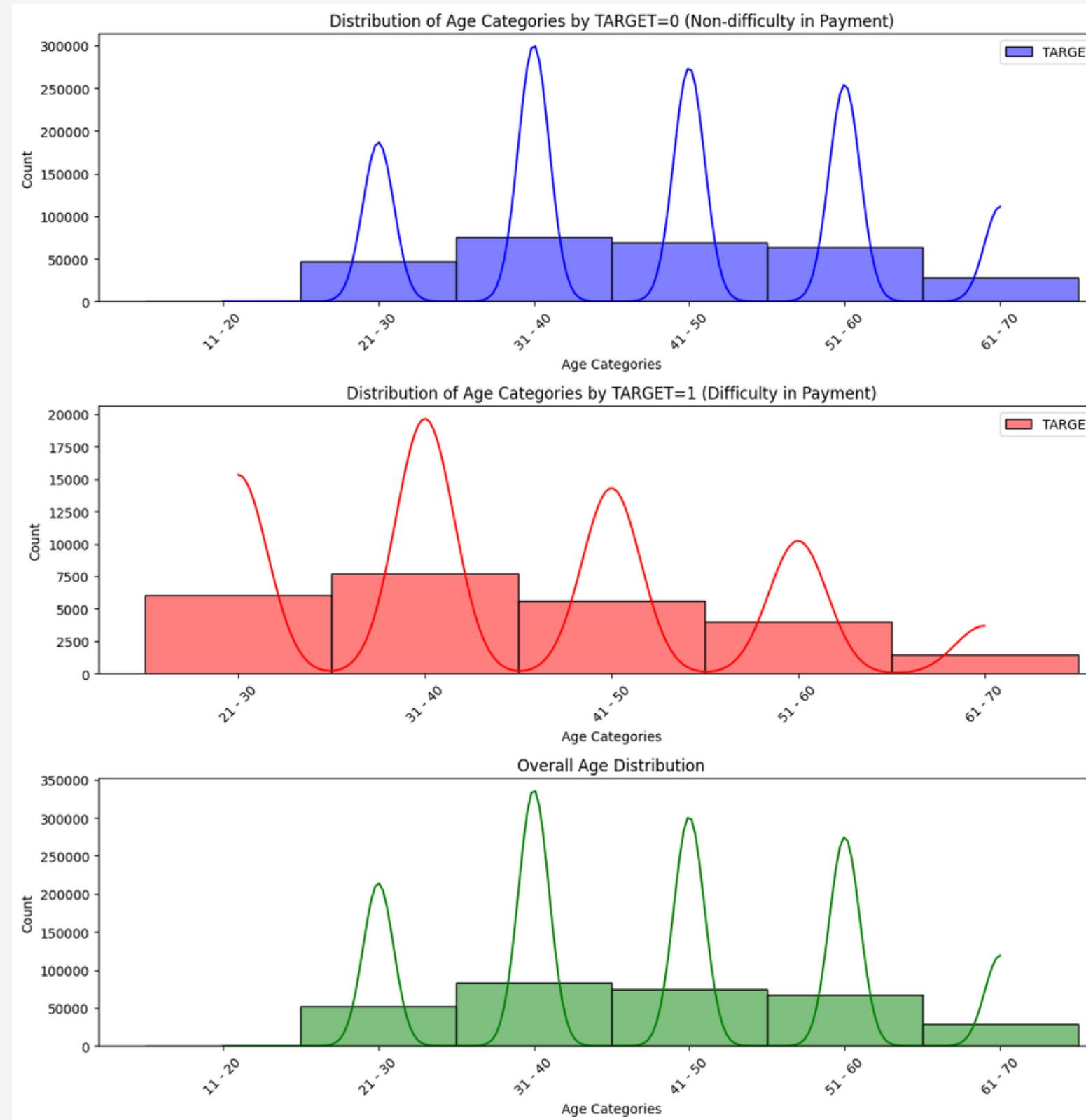
### Splitting Data

Data is split into **data train** and **data test** with **80 : 20 ratio**

### Transformation Feature & Handling Class Imbalance

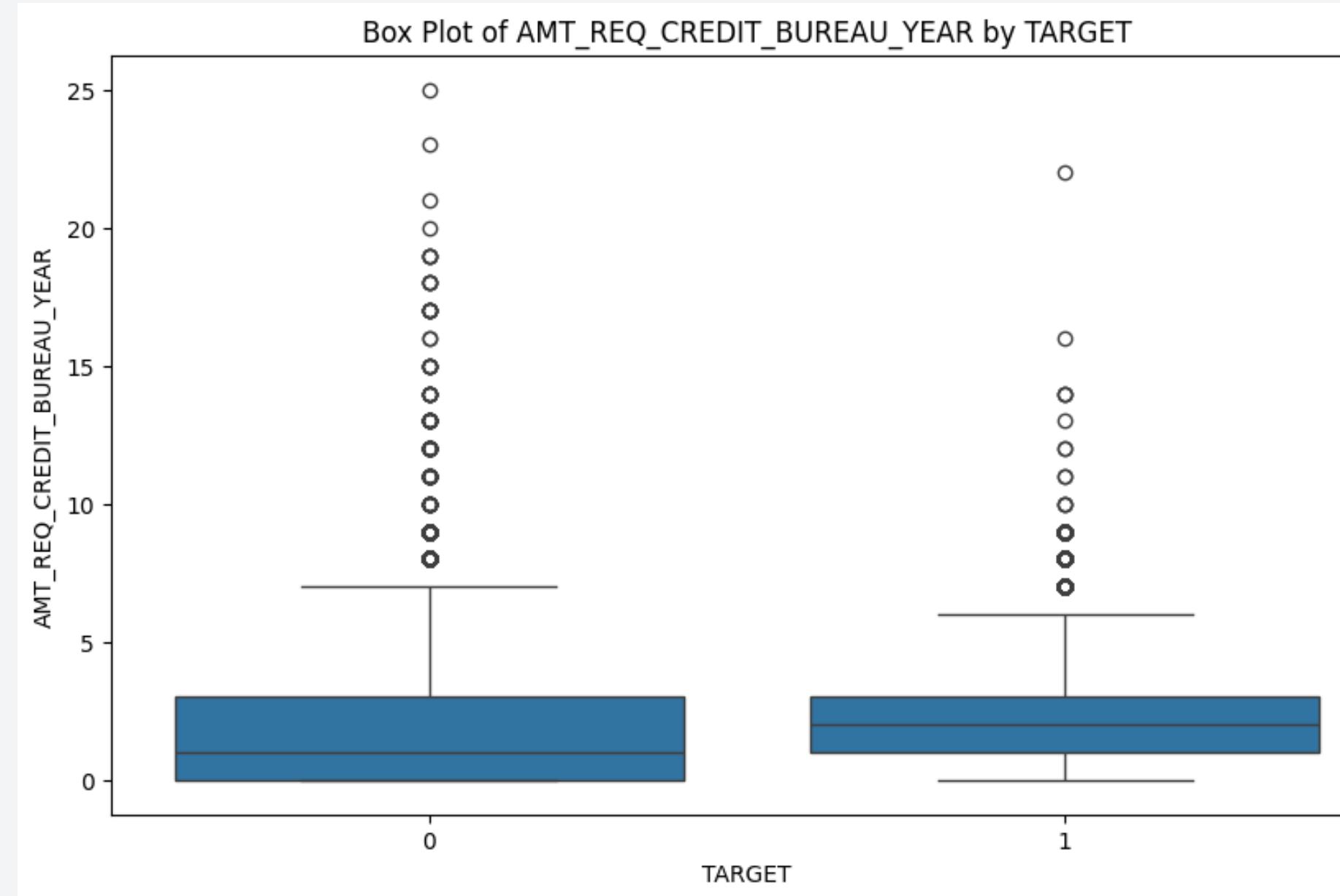
- **Standard scalar** used for **feature transformation** due to non-normal distribution data
- **SMOTE (oversampling)** used for handling class imbalance

# 3.) DATA VISUALIZATION AND BUSINESS INSIGHT



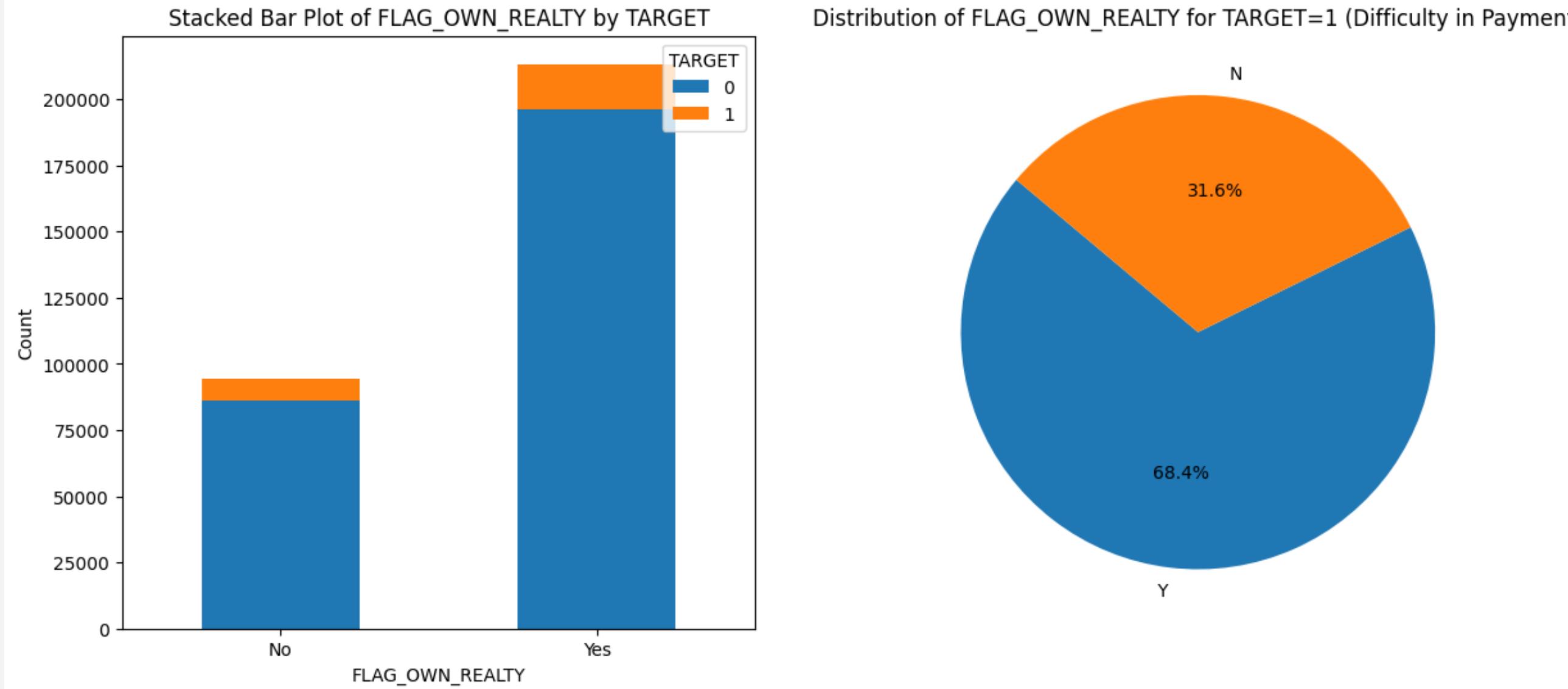
- The majority of loan applicants fall within the 31-40 age bracket, while the number of applicants < 31 or > 60 years old is notably minimal.
- Applicants who experience no payment issues typically range from 61 to 70 years old. Conversely, those encountering payment difficulties generally belong to the 31-40 age group.

### 3.) DATA VISUALIZATION AND BUSINESS INSIGHT



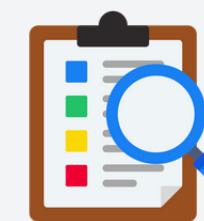
Individuals with more inquiries to the Credit Bureau tend to experience difficulty in repaying loans. This may indicate they are seeking various types of loans or have a complex credit history.

### 3.) DATA VISUALIZATION AND BUSINESS INSIGHT



Individuals who own a house or flat tend to have difficulty repaying loans. This is interesting and require more in depth exploration by the company in order to find the “WHY”.

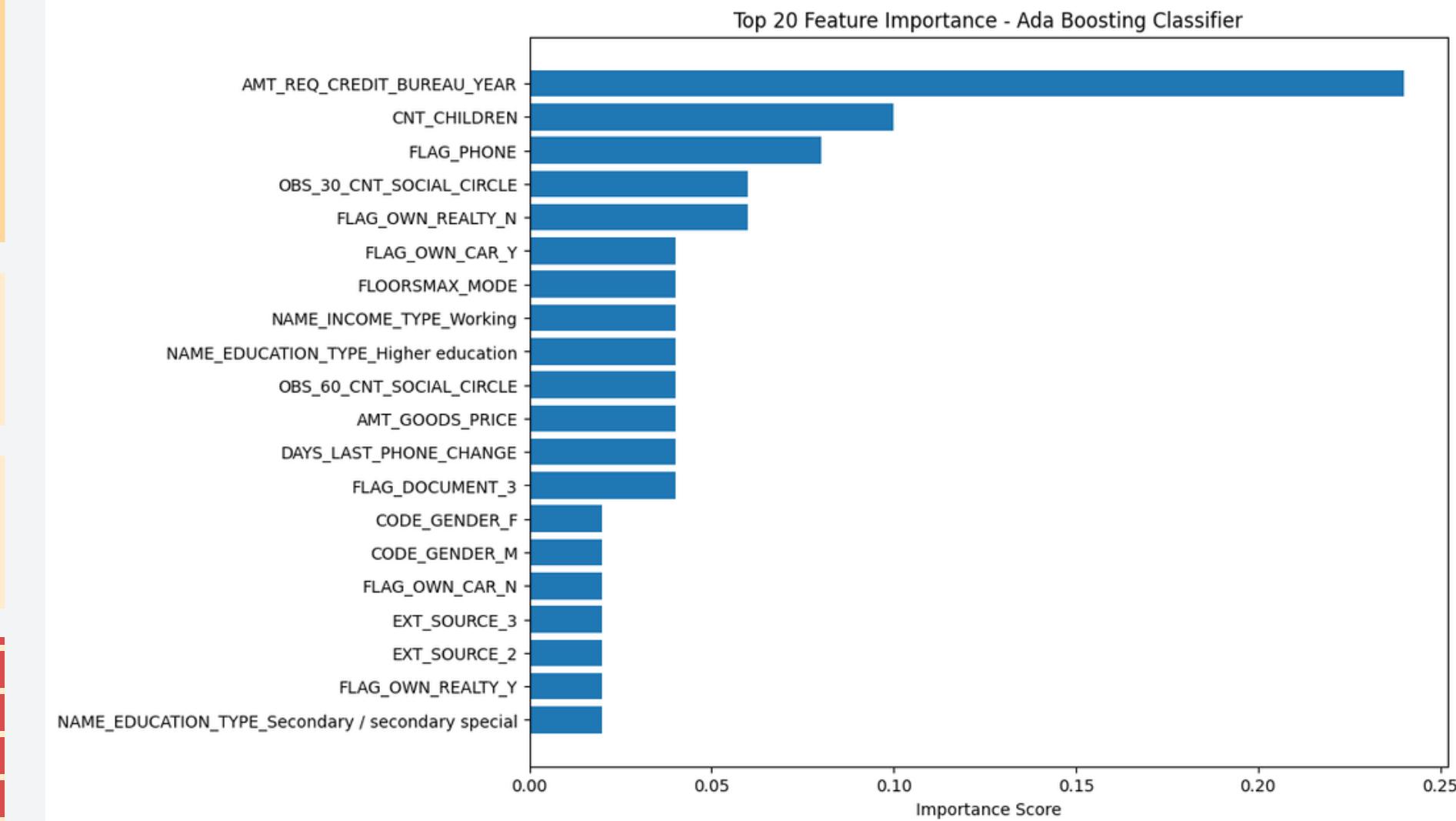
# 4.) MACHINE LEARNING IMPLEMENTATION AND EVALUATION



## Model Evaluation - Comparison

Models	Precision (in %)		Recall (in %)		ROC AUC (in %)
	Training	Testing	Training	Testing	
Logistic Regression	70.00	70.00	70.00	70.00	70.25
Decision Tree Classifier	100.00	87.00	100.00	87.00	87.15
<b>Ada Boosting Classifier</b>	89.00	90.00	89.00	89.00	89.49

## Features Importance



**Ada Boosting Classifier** has been selected as the optimal model based on the consistency observed in training and testing results, where significant discrepancies were not evident.

- Logistic Regression (Underfitting)
- Decision Tree Classifier (Overfitting)

## Top 5 Important Features

1. AMT\_REQ\_CREDIT\_BUREAU\_YEAR (**0.24**)
2. CNT\_CHILDREN (**0.10**)
3. FLAG\_PHONE (**0.08**)
4. OBS\_30\_CNT\_SOCIAL\_CIRCLE (**0.06**)
5. FLAG\_OWN\_REALTY\_N (**0.06**)

## 5.) BUSINESS RECOMMENDATION

01

### Referring to AMT\_REQ\_CREDIT\_BUREAU\_YEAR (Number of enquiries to Credit Bureau per year)

The company can pay special attention to customers with a high number of inquiries to the Credit Bureau, perhaps by applying stricter risk assessments or offering financial solutions tailored to their needs.

02

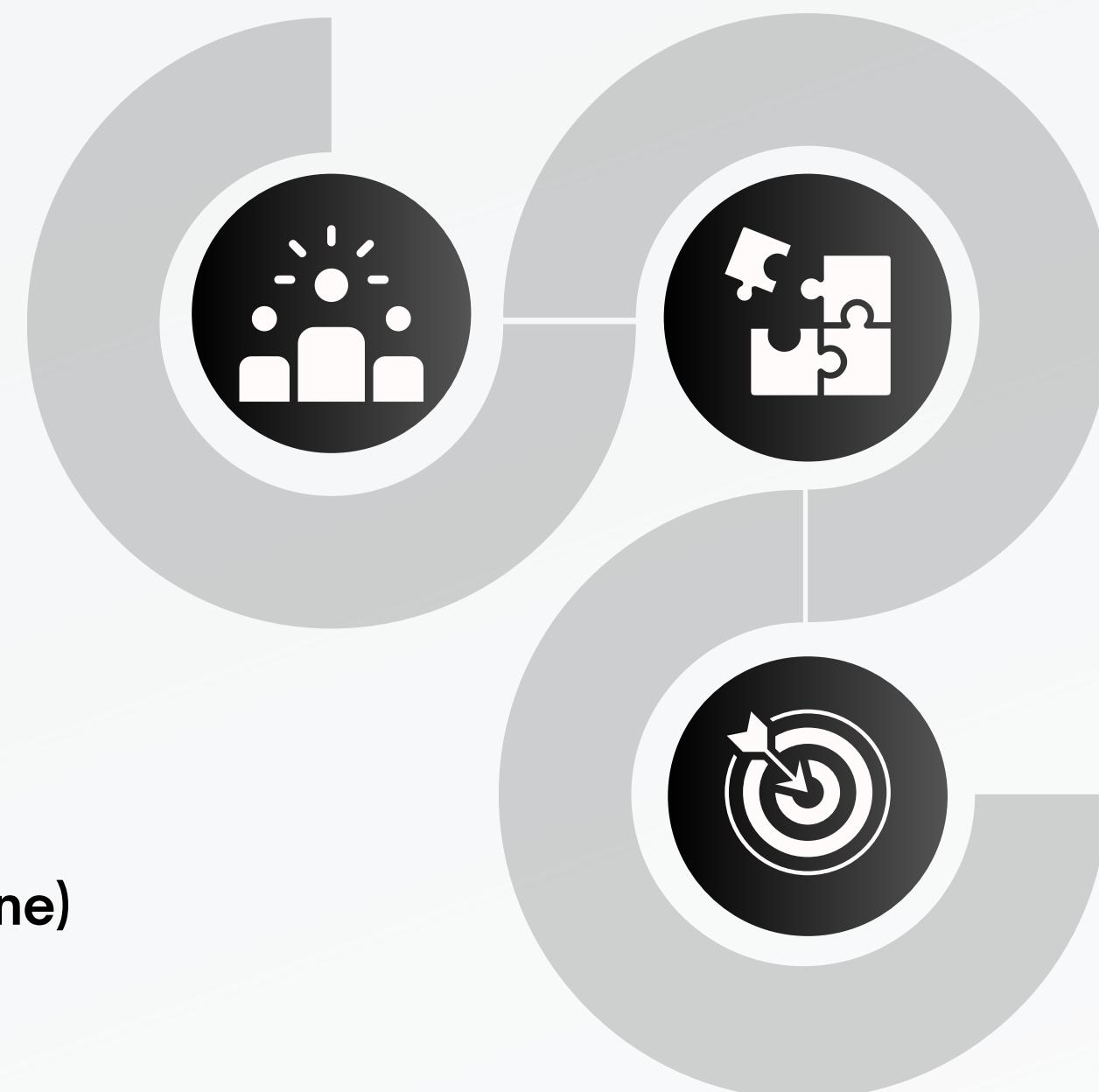
### Referring to CNT\_CHILDREN (Number of children)

The company can tailor its products or services, or provide programs or benefits specifically for individuals with dependent children.

03

### Referring to FLAG\_PHONE (Availability of home phone)

The company could consider alternative communication methods or more personal approach strategies for customers who do not provide a home phone number.



## 5.) BUSINESS RECOMMENDATION

04

**Referring to OBS\_30\_CNT\_SOCIAL\_CIRCLE (Number of observations from social surroundings in 30 days)**

The company may consider using additional information about clients' social environment or networks to enrich risk assessment and decision-making.

05

**Referring to FLAG\_OWN\_REALTY\_N  
(Does not own a house or flat)**

The company can adjust its risk assessment strategy to consider clients' homeownership status and develop products or services more aligned with their needs.



# THANK YOU

