

Muhammad Muhibuddin Mukhlish



CREDIT RISK ANALYSIS

PREDICTING LOAN DEFAULT USING
MACHINE LEARNING

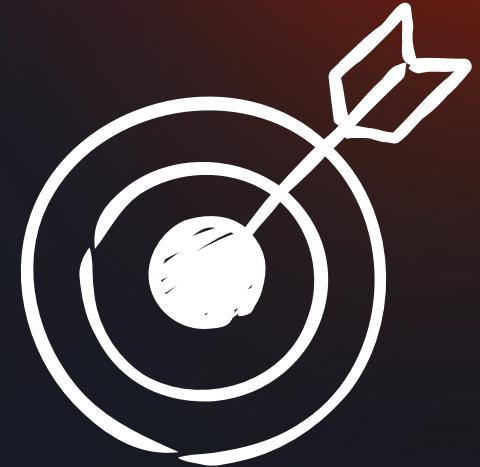
VIX - VIRTUAL INTERNSHIP EXPERIENCE

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 the number of customers authorized for approval but who ultimately default.



OBJECTIVE

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 EVALUTION

Precision of the model in predicting customer defaulters on a loan.

DATA PRE-PROCESSING

1

HANDLING MISSING VALUE

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.

2

FEATURE ENGINEERING & SELECTION

Applied One-Hot Encoding to 13 categorical columns. Dropped 'SK_ID_CURR' and redundant 'AMT_REQ_CREDIT_BUREAU' columns (Hour, Day, Week, Mon, Qrt).

3

TRANSFORMATION & HANDLING CLASS IMBALANCE

Normalized numerical features using StandardScaler. Handled imbalance using SMOTE (Oversampling) on the training data.

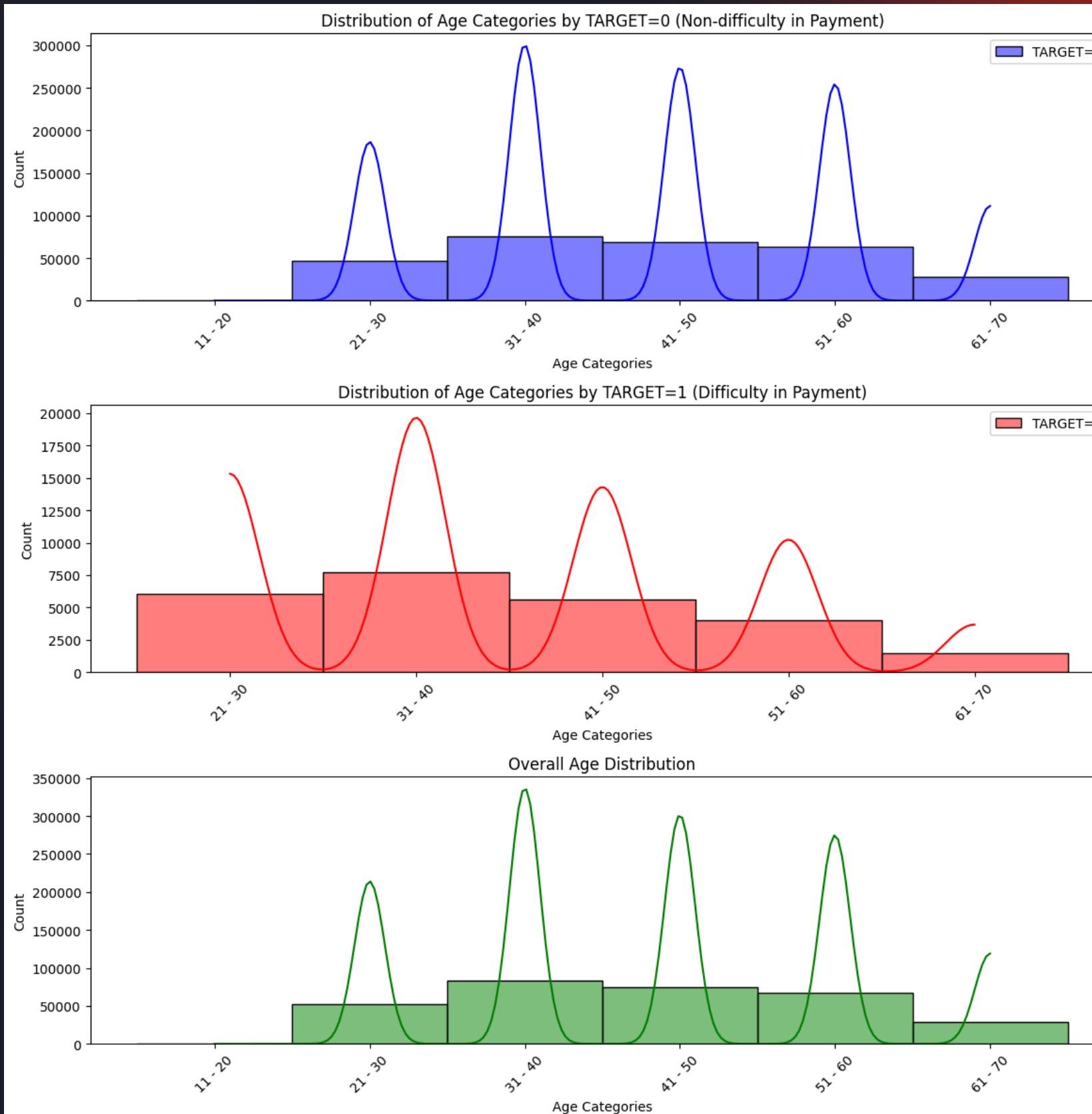
4

SPLITTING DATA

Data was split into an 80% train and 20% test set.

DATA VISUALIZATION AND BUSINESS INSIGHT

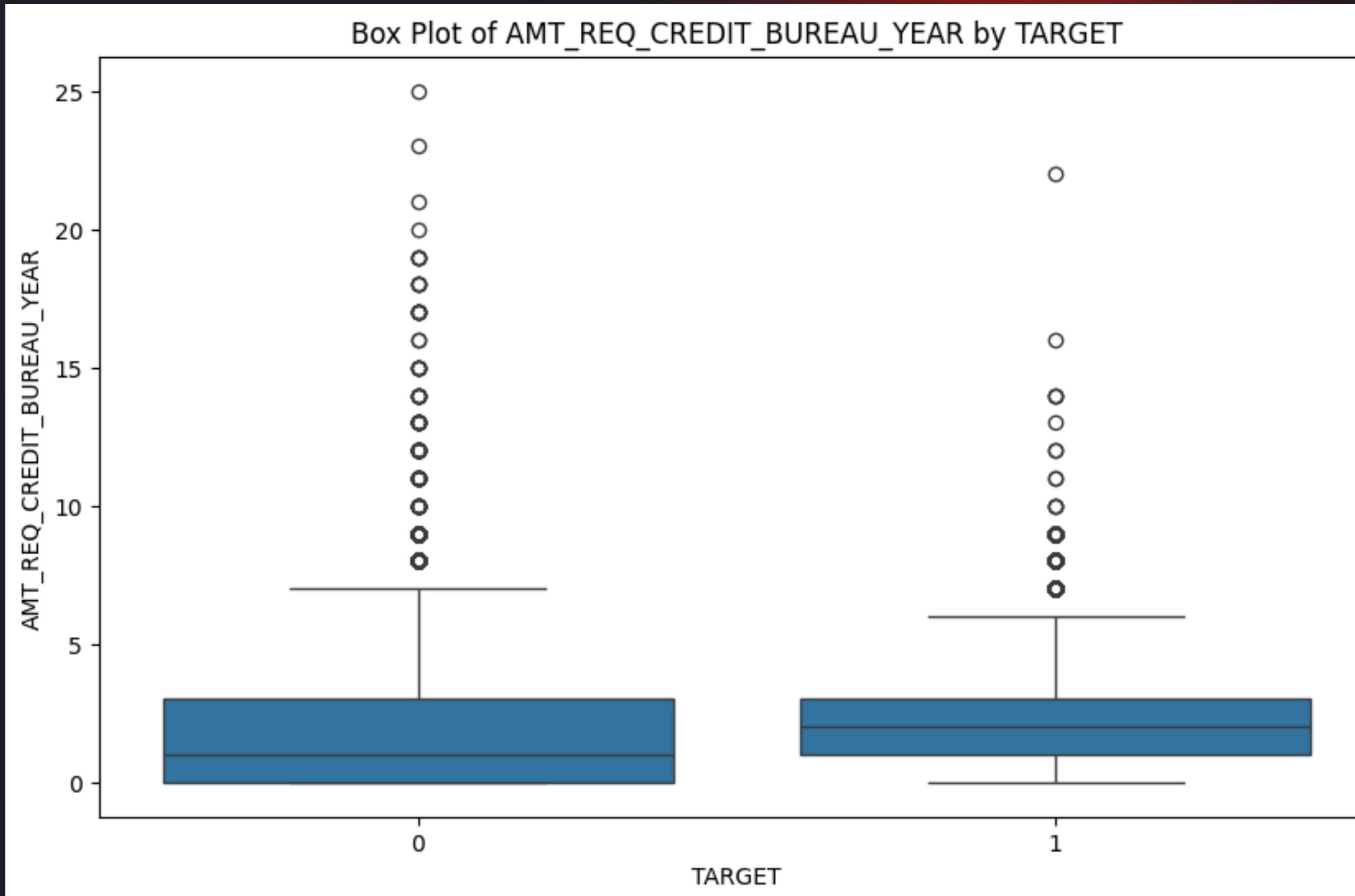
Muhammad Muhibuddin Mukhlish



- 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.

DATA VISUALIZATION AND BUSINESS INSIGHT

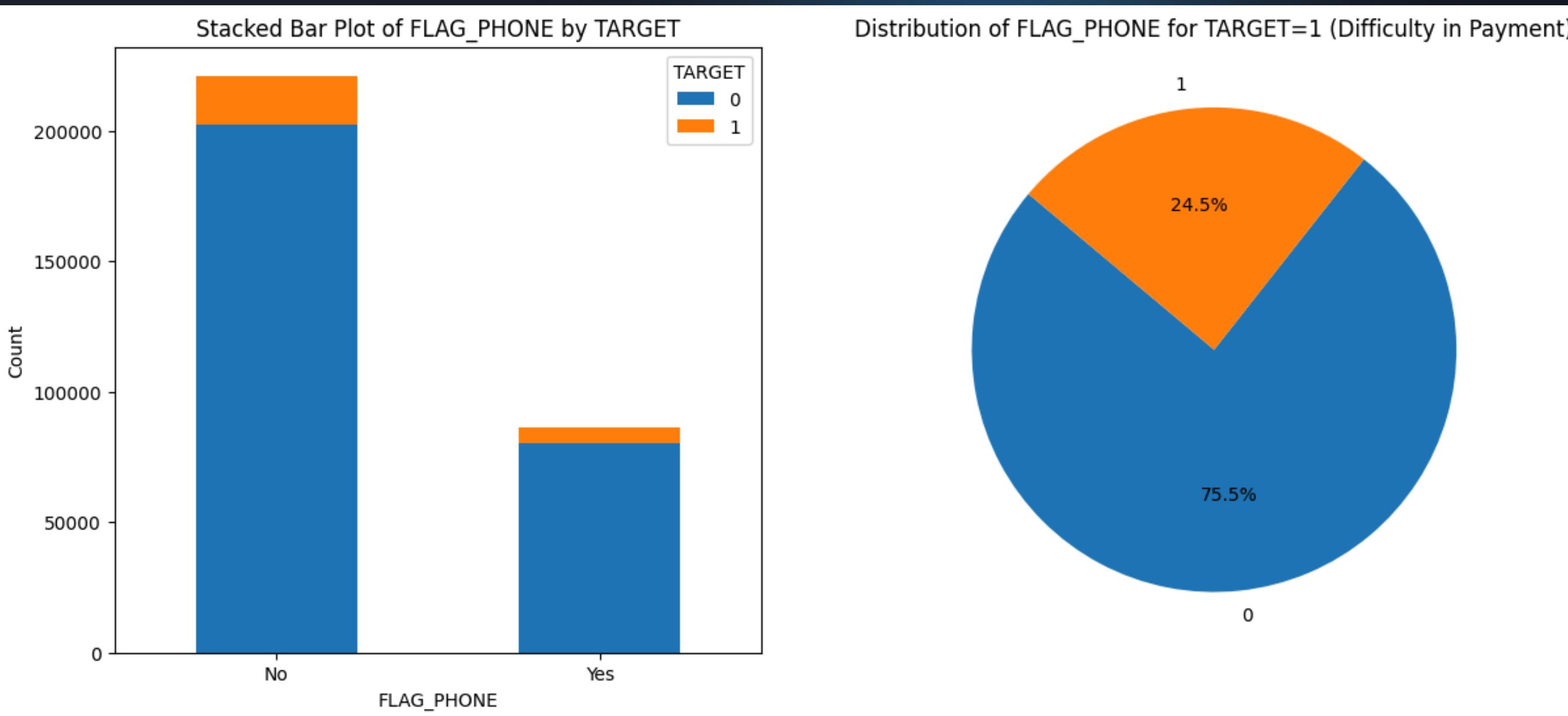
Muhammad Muhibuddin Mukhlish



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.

DATA VISUALIZATION AND BUSINESS INSIGHT

Muhammad Muhibuddin Mukhlish



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"

MACHINE LEARNING IMPLEMENTATION AND EVALUATION

Muhammad Muhibuddin Mukhlish

Models	Precision (in %)	Recall (in %)	ROC AUC (in %)
	Training	Testing	Training
Logistic Regression	70	70	70
Decision Tree Classifier	100	87	100
Ada Boosting Classifier	86	86	86

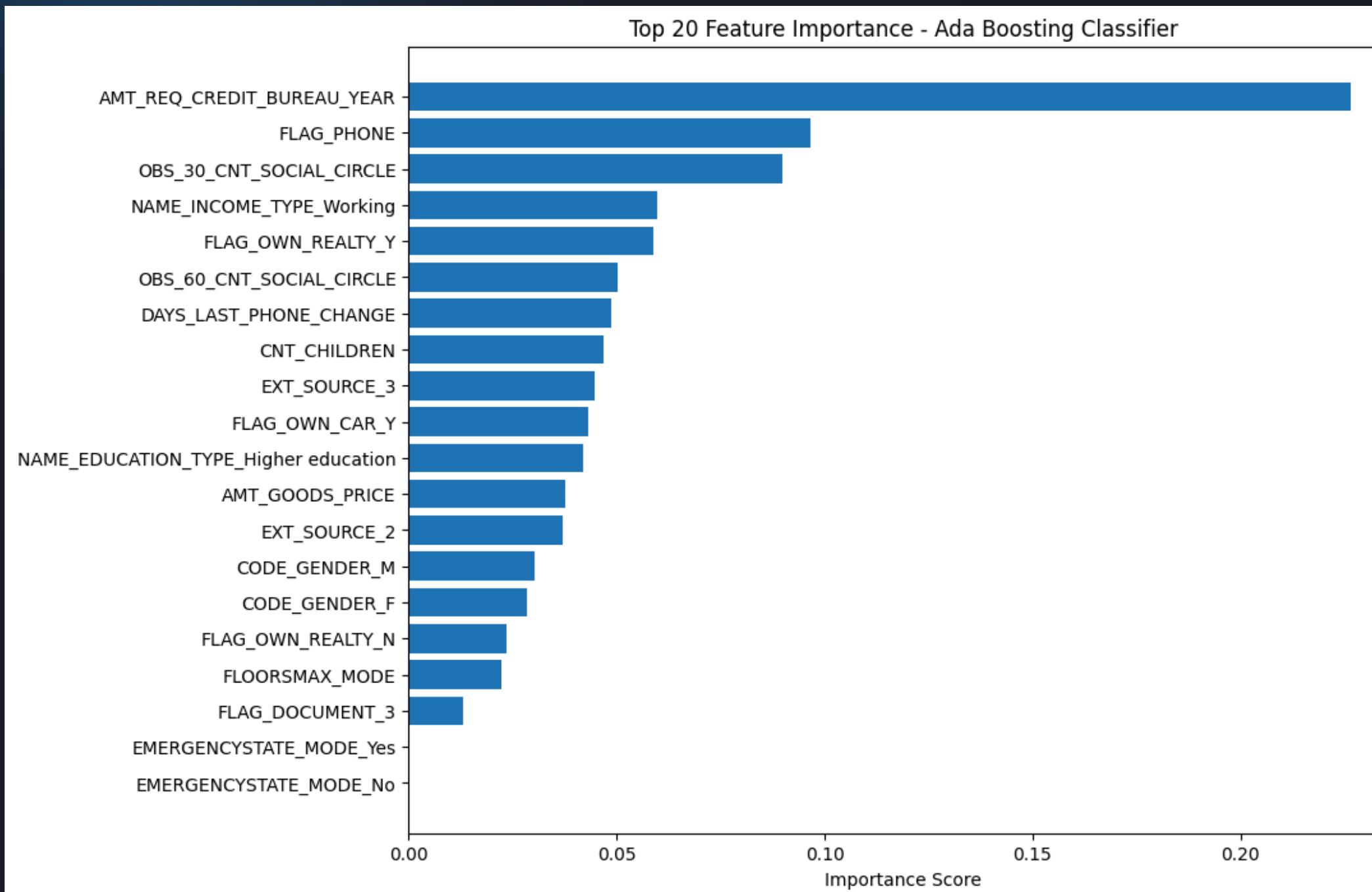
Model Evaluation - Comparison

Ada Boosting Classifier has been selected as the optimal model. It shows the best and most consistent performance between the training results (85.85% Accuracy) and the testing results (86.50% Accuracy), indicating it generalizes well.

- Logistic Regression (Underfitting): This model performed poorly on both training (70.22% Acc) and testing data (69.96% Acc).
- Decision Tree Classifier (Overfitting): This model achieved a perfect score on training data (100.0% Acc) but showed a significant drop in performance on test data (87.35% Acc).

MACHINE LEARNING IMPLEMENTATION AND EVALUATION

Muhammad Muhibuddin Mukhlish



Features Importance

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)

BUSINESS RECOMMENDATION



Muhammad Muhibuddin Mukhlish

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.

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.

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.

BUSINESS RECOMMENDATION

Muhammad Muhibuddin Mukhlish

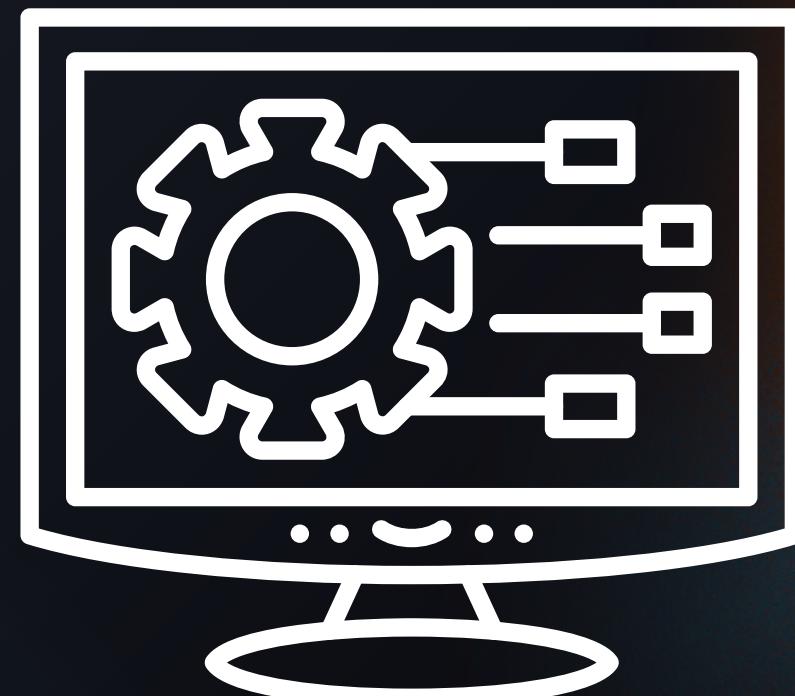
Referring to OBS_30_CNT_SOCIAL_CIRCLE (Number of observations from social surroundings in 30 days)

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

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.

#e21932



Muhammad Muhibuddin Mukhlish

THANK YOU

