

Bureau Loan Prediction Assignment Report

Piyush Mohapatra
piyushm.ofcl@gmail.com

August 18, 2024

1 Introduction

This report presents the approach taken to solve the Bureau Loan Prediction assignment. The goal was to predict whether a two-wheeler loan application would be accepted or rejected based on the provided data. I employed a Random Forest Classifier after appropriate preprocessing of the data. This report details the approach, insights from the data, and performance metrics of the model.

2 Approach Taken

2.1 Data Preprocessing

- **Handling Missing Values:** Missing numerical values were filled using the mean, and missing categorical values were filled using the most frequent value. This was done using `SimpleImputer`.
- **Categorical Encoding:** I used `OrdinalEncoder` with `handle_unknown='use_encoded_value'` to handle unseen labels by assigning them a value of -1.
- **Feature Scaling:** Numerical features were scaled using `StandardScaler` to ensure consistency across features, although this step is optional for tree-based models.
- **Model Selection:** A Random Forest Classifier was trained on the processed data due to its ability to handle complex feature interactions and its robustness against overfitting.

3 Insights and Conclusions from Data

- **Imbalanced Data:** The dataset showed a slight imbalance between approved and declined applications, with more approvals than declines. Despite this, the model performed well on both classes.

- **Feature Importance:** Key features such as `Cibil Score`, `Total Asset Cost`, and `Applied Amount` were found to be highly important in predicting loan approval status.
- **Model Performance:** The model achieved a validation accuracy of 88.6%, with strong precision and recall metrics, particularly for the approved class.

4 Performance on Train Data Set

- **Accuracy:** 88.6%
- **Precision:**
 - Approved: 90%
 - Declined: 85%
- **Recall:**
 - Approved: 93%
 - Declined: 80%
- **F1-Score:**
 - Approved: 92%
 - Declined: 82%

The Random Forest model showed a balanced performance between approved and declined classes, with slightly better results on the approved class.

5 Code

- **Code Repository:** The complete code, including data preprocessing, model training, and prediction generation, has been uploaded to GitHub. The repository can be accessed at:
 - <https://github.com/piyush-mk/Bureau-Loan>
- **Code Notebook:** The Jupyter notebook containing the code can be accessed directly at:
 - <https://github.com/piyush-mk/Bureau-Loan/blob/main/Code.ipynb>