Loan Prediction Model Report

Introduction

The objective of this project is to develop a machine learning model that can accurately predict loan defaults. Utilizing a dataset containing various features such as disbursement amount, asset cost, and employment type, a Random Forest Classifier was employed for this predictive task.

Data Preprocessing

- Missing values were filled with 'Unknown' and then label-encoded.
- Date columns were dropped after relevant features were engineered.
- The features 'disbursed_amount' and 'asset_cost' were scaled using Standard Scaler.

Model Performance

The model achieved an accuracy of approximately **77.6%**. The Confusion Matrix and Classification Report were as follows:

Confusion Matrix:

[[35641, 836],

[9611, 543]]

Classification Report:

- Precision: 0.79 (for class 0), 0.39 (for class 1)

- Recall: 0.98 (for class 0), 0.05 (for class 1)

Cross-Validation

5-fold cross-validation was employed, yielding scores of `**[0.78076816, 0.77476357, 0.76174648, 0.76629281, 0.76056187]**`. The average score is around **76.9%**, indicating a robust model with room for improvement.

Conclusion

The model provides a reasonably accurate and consistent prediction of loan defaults. However, its performance on minority class (loan default = 1) is low, requiring further improvement for practical deployment.