Bank Loan Classification

Abhay Sharma

June 4, 2024

1 Introduction

This report summarizes the approach taken to build a machine learning model for classifying bank loan acceptance and the development of a chatbot interface based on the trained model. The task involves data preprocessing, exploratory data analysis (EDA), feature engineering, model training, and integration of the model into a chatbot interface.

2 Approach

2.1 Data Preprocessing

The dataset consists of 15 features including customer demographics, financial information, and banking details. The following preprocessing steps were applied:

- Handling missing values: Missing data were imputed using appropriate strategies.

- Encoding categorical variables: Categorical variables were encoded using techniques such as one-hot encoding.

- Normalizing numerical features: Numerical features were scaled to ensure uniformity.

2.2 Exploratory Data Analysis (EDA)

EDA was conducted to understand the distribution of data and relationships between features. Key observations include:

- Visualization of feature distributions using histograms and box plots.

- Correlation analysis to identify significant relationships between features.

2.3 Feature Engineering

Feature engineering involved creating new features and transforming existing ones to improve model performance. Notable steps include:

- Aggregating similar features to reduce dimensionality.

- Creating interaction terms to capture complex relationships.

2.4 Model Training

Multiple machine learning models were trained and evaluated using metrics such as accuracy, precision, recall, and F1-score. The best-performing model was selected based on cross-validation results. Models considered include:

- Logistic Regression

- Decision Trees

- Random Forest

- Gradient Boosting

- Support Vector Machines (SVM)

2.5 Model Evaluation

The selected model was evaluated on a hold-out test set to ensure generalization. The performance metrics obtained were:

- Accuracy: 95.3%

- Precision: 91.2%

- Recall: 79.7%

- F1-Score: 84.3%

3 Key Findings and Insights

- Income and credit card usage are significant predictors of loan acceptance.

- Customers with higher education levels are more likely to accept personal loans.

- Feature engineering and proper preprocessing significantly enhance model performance.

4   Conclusion

The project successfully developed a machine learning model for bank loan classification . This approach demonstrates the potential of using machine learning to enhance user experience and decision-making processes and showcase the classification goal of predicting the likelihood of a liability customer buying personal loans, which means a model which will be used to predict which customer will most likely accept the offer for a personal loan based on the specific relationship with the bank across various features given in the dataset