Summary Report: Bank Loan Prediction

Introduction:

The purpose of this analysis is to predict whether a customer will accept a personal loan based on various features such as age, income, education, home ownership, and more. The dataset contains information on customers, including their demographics, financial attributes, and loan acceptance status.

Data Exploration:

The dataset consists of 15 features, including ID, age, gender, experience, income, home ownership, zip code, family size, average monthly credit card spending, education level, mortgage value, securities account, CD account, online banking usage, credit card usage, and personal loan acceptance.

Data Preprocessing:

Missing Values: Missing values are handled by filling them with appropriate values. Numeric columns are filled with the mean, and non-numeric columns are filled with the most frequent value.

Data Scaling: Numeric features are normalized using Min-Max scaling to ensure they have the same scale.

Data Analysis and Visualization:

Loan Acceptance Distribution: A count plot is used to visualize the distribution of loan acceptance, showing the number of customers who accepted or rejected the personal loan.

Income Distribution by Loan Acceptance: A boxplot is created to compare the income distribution for customers who accepted and rejected the personal loan.

Correlation Analysis: A correlation matrix and heatmap are generated to explore the relationships between the features and identify any significant correlations.

Model Training and Evaluation:

The dataset is split into training and testing sets.

Three classifiers (Logistic Regression, SVM, Random Forest) are trained on the training set and evaluated on the testing set using accuracy as the performance metric.

The Random Forest classifier performs the best among the three models, achieving the highest accuracy on the test set.

Conclusion:

The analysis demonstrates that it is possible to predict personal loan acceptance based on customer features. The Random Forest classifier is recommended for making predictions with the highest accuracy. Further optimization and fine-tuning of the model can be explored to improve prediction performance.