Title: Analysis of Bank Loan Acceptance

Date: 3rd August 2023

Introduction:

In this report, I present the analysis of a dataset containing information about bank loans, aiming to classify whether a personal loan was accepted or not based on the provided data.

Approach:

1. Data Preprocessing:

- o Loaded the dataset from the provided Excel file.
- o Handled missing values by imputing or replacing them, ensuring data integrity.
- o Converted categorical variables into numerical format using label encoding.
- o Normalized numerical features to bring them to a similar scale.

2. Exploratory Data Analysis (EDA):

- o Explored the distribution of the target variable ('Personal Loan') to understand class balance.
- o Analyzed numerical feature distributions and identified potential outliers.
- o Visualized relationships between features and the target variable using plots.

3. Feature Engineering:

- o Converted categorical variables into numerical format using label encoding.
- o Normalized numerical features to bring them to a similar scale.

4. Model Selection and Training:

- o Split the data into training and testing sets.
- o Chose the Logistic Regression as the model for its ability to handle categorical targeted variables.
- o Trained the model using the preprocessed data.
- o Evaluated model performance on the testing data using classification metrics.

Key Findings:

1. The distribution of the target variable showed an imbalance, with fewer accepted personal loans.
2. Feature importance analysis indicated that 'Income', 'CCAvg', and 'CD Account' were significant predictors of loan acceptance.

Insights and Observations:

1. Clients with higher incomes and credit card spending were more likely to accept personal loans.
2. Holding a certificate of deposit (CD) account seemed to increase the likelihood of accepting a loan.
3. The class imbalance in the target variable might impact model performance.