Problem statement

You are given a dataset of the bank loans consisting of 15 columns and a corresponding target column. Your task is to build a machine-learning model that can accurately classify
whether the personal loan was accepted or not based on the information provided.

Methodology

Below are the method that I have used to built my model:

Data preprocessing

- simple imputers is used to fill the missing value with mean of that features.
- one hot encoder is used to convert the categorical features to one hot encoded value.
- Standard scaler is used to normalized the value of features so that all the features will give equal contribution to predict the target.

Feature selection

Selected the features like 'Age', 'Experience', 'Income', 'Family', 'CCAvg', 'Education', 'Mortgage', 'Securities Account', 'CD Account', 'Online', 'Credit Card', 'Gender', 'Home Ownership' except zip code because the location wont affect the personal loan approval.

Data Split

I have spitted the data into training dataset to train my model and testing dataset to test the model. The ration of training and testing dataset is 4:1

Model Selection

I have selected Logistic Regression to train my dataset because the logistic regression uses sigmoid function which will distinguish two different classes.

Accuracy Calculation

Accuracy Score is used to calculated the accuracy of my model which is compare predicted value and actual value.