ML Lab Assignment 3 – Part 2

Q1 . Using the Naive Bayes Classifiers classify potential customers who are more likely to purchase a loan by the Bank_Personal_Loan_Modelling dataset. The dataset description is given below.

Dataset:
The Bank_Personal_Loan_Modelling dataset includes 5000 observations with fourteen variables divided into four different measurement categories. The binary category has five variables, including the target variable personal loan, also securities account, CD account, online banking and credit card. The interval category contains five variables: age, experience, income, CC avg and mortgage. The ordinal category includes the variables family and education. The last category is nominal with ID and Zip code. The variable ID and Zip code does not add any interesting information . E.g. individual association between a person (indicated by ID) and loan does not provide any general conclusion for future potential loan customers. Therefore, it will be neglected in the examination.

Q2. Do the performance analysis of the above model using precision, recall and F1-score.

Q3. Plot the distribution of the input features.
[Note: In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution. When plotted, it gives a bell shaped curve which is symmetric about the mean of the feature values.]

Q4. Using this Bayesian classifier, make a function which takes input as Age, Income and Experience . Output: Predict whether loan will be given or not.