

# LAB REPORT -4

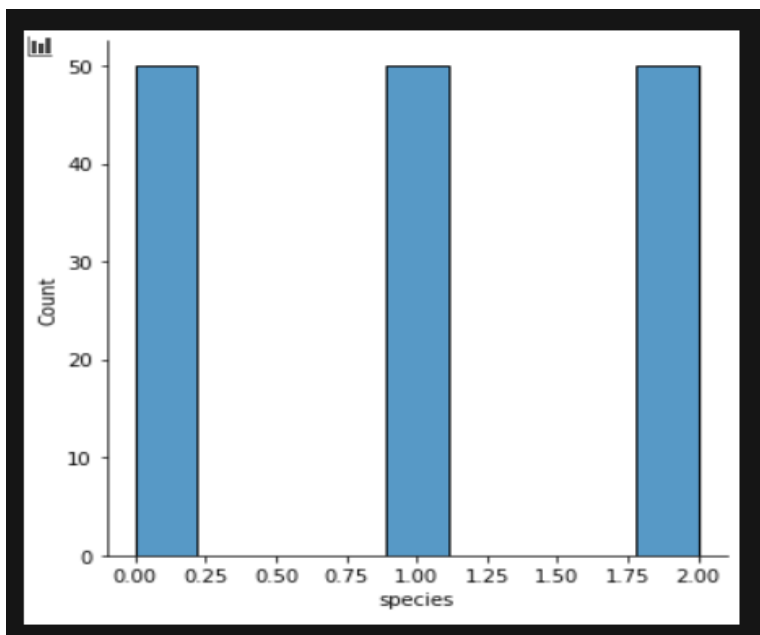
**Mayank Raj (B19CSE053)**

Btech Cse

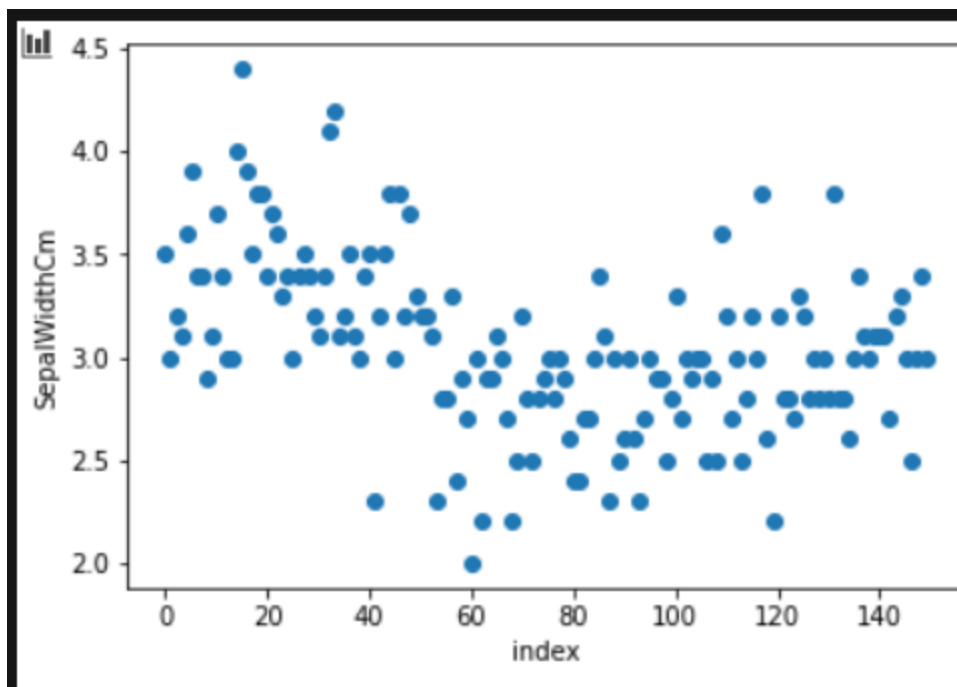
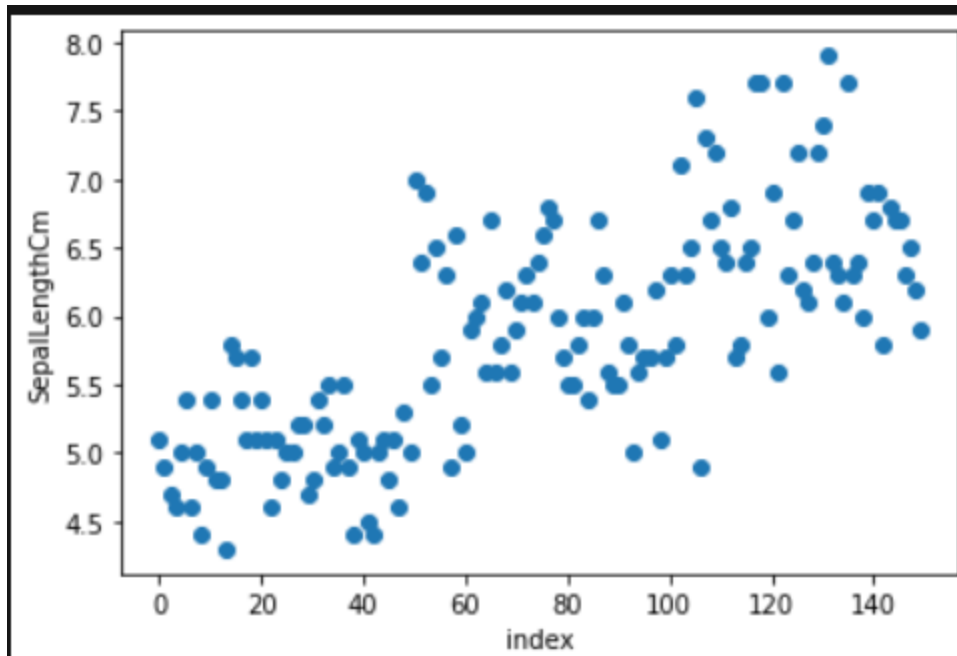
Q1)

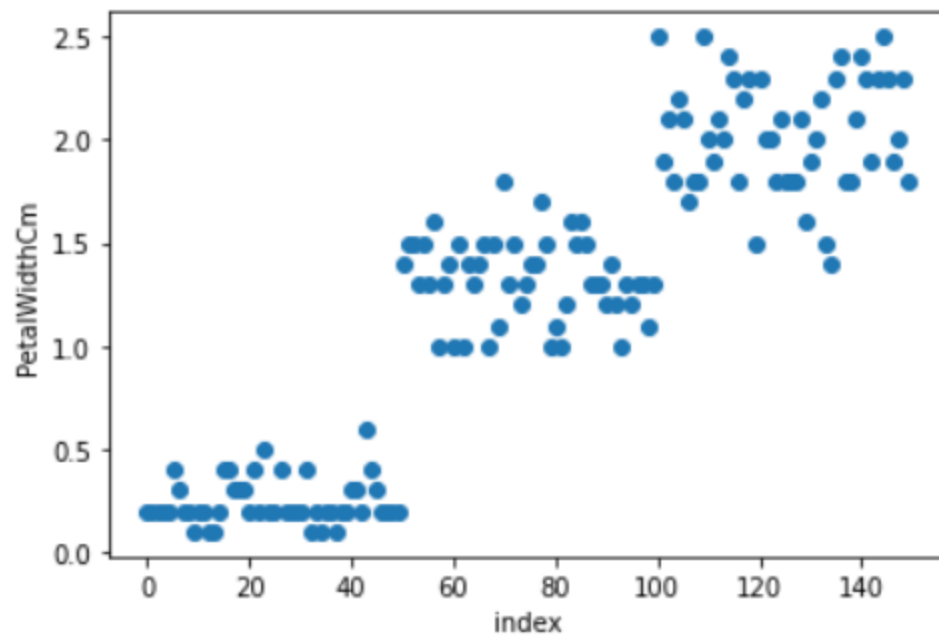
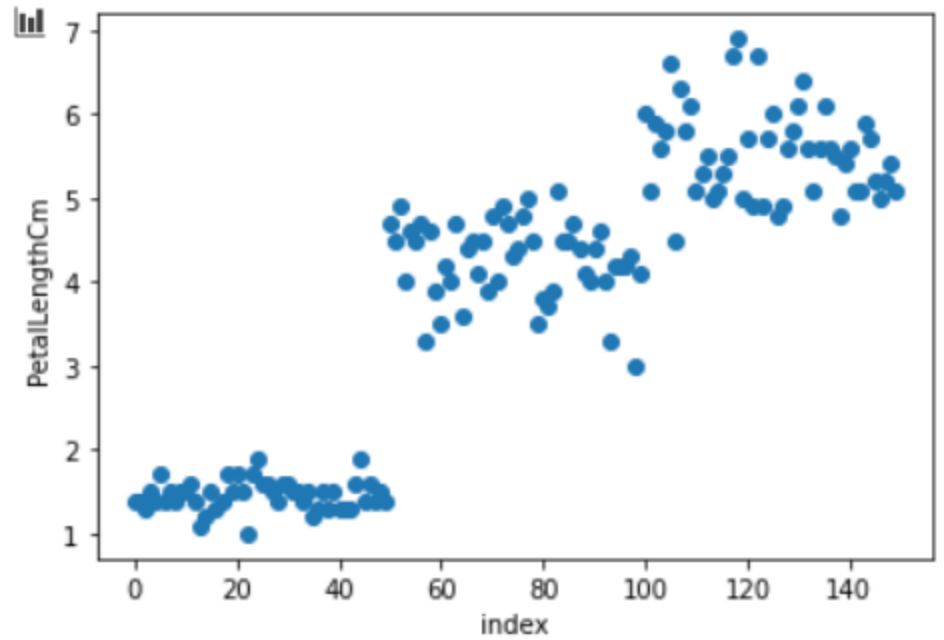
1)Preprocessing:

- Plotted the distribution of the target variable



- Visualizing the distribution of data for every feature.



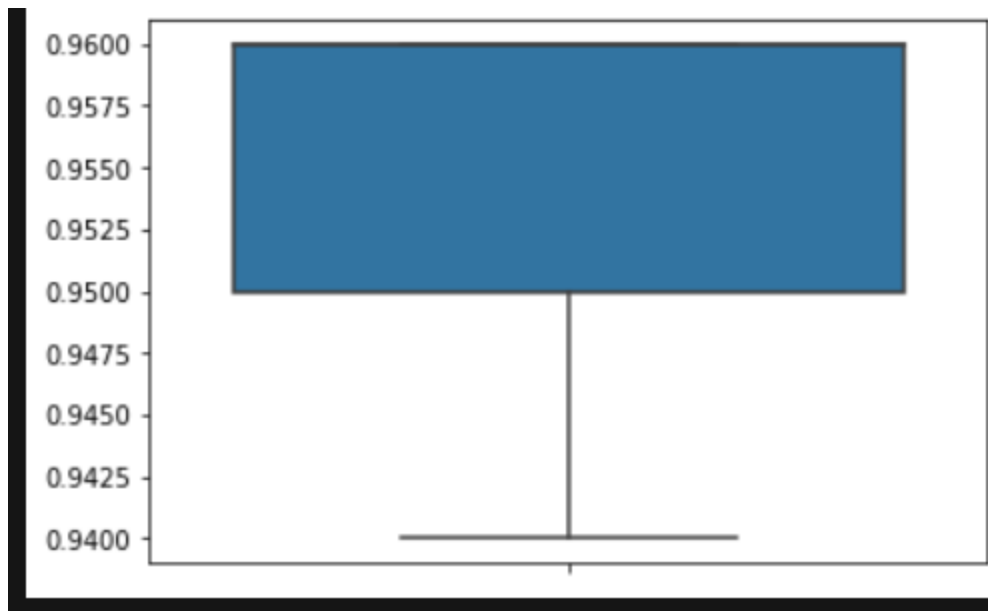


2) Performed boosting-based classification using Decision Tree as the base classifier

- Boosting is performed with decision tree as the base classifier with max\_depth = 2 and accuracy achieved by AdaBoost was 94-96%

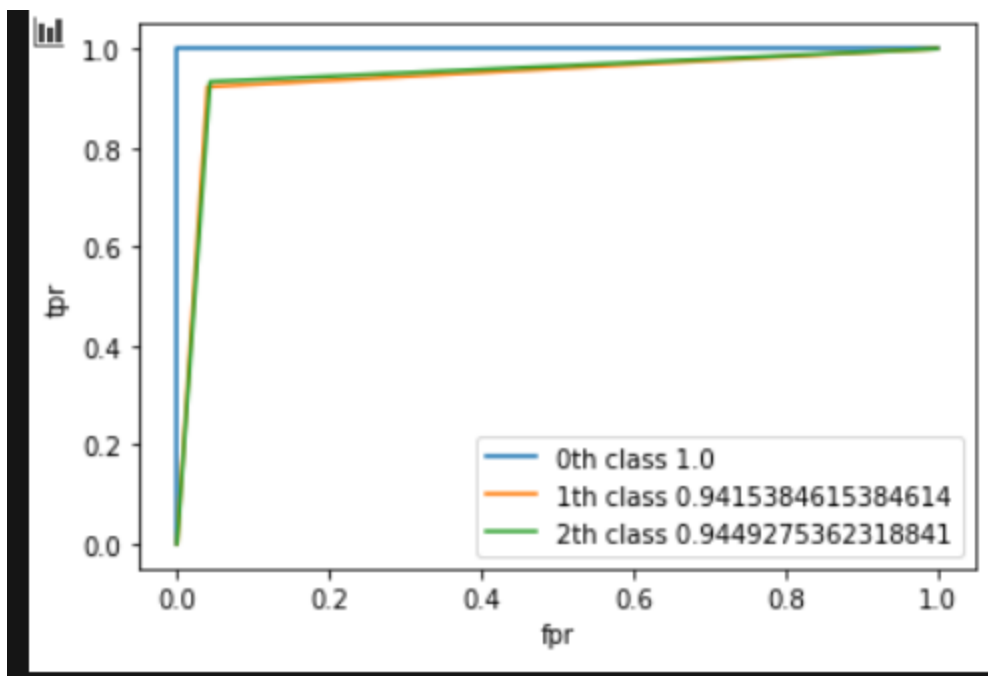
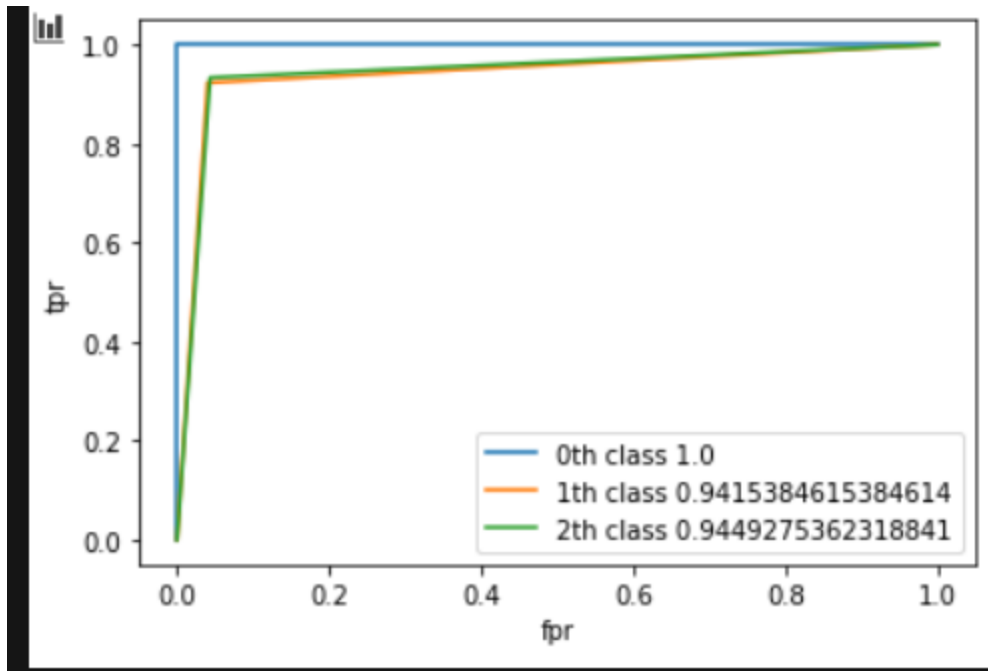
3) Performed cross validation over the data and calculate accuracy for a weak learner

- 3 fold cross validation was done and cross validation score of mean 0.95333 was achieved.



4) Built the AdaBoost model using the weak learner by increasing the number of trees from 1 to 5 with a step of 1. Compute the model performance

- Estimators range was taken from 10 to 100 with increase in 10

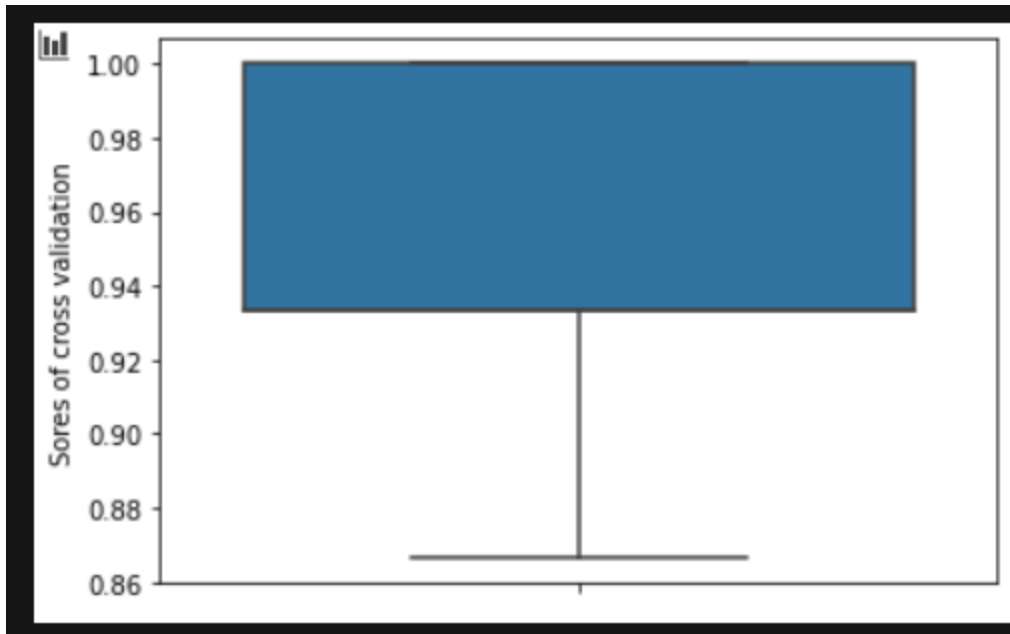


The Adaboost classifier gives accuracy of 96-100% in this case.

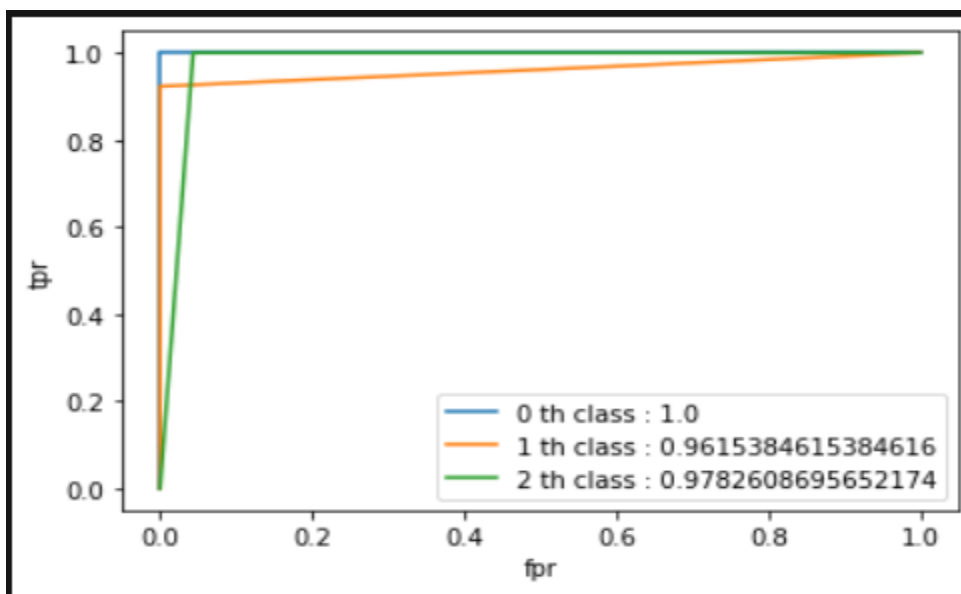
Q2)

1) Estimated the accuracy of Naive Bayes algorithm using 5-fold cross validation on the data set. Plot the ROC AUC curve for different values of parameters

- 10 fold cross validation was done on the naive Bayes model and mean accuracy of 95.333 was achieved.



- Roc curve:



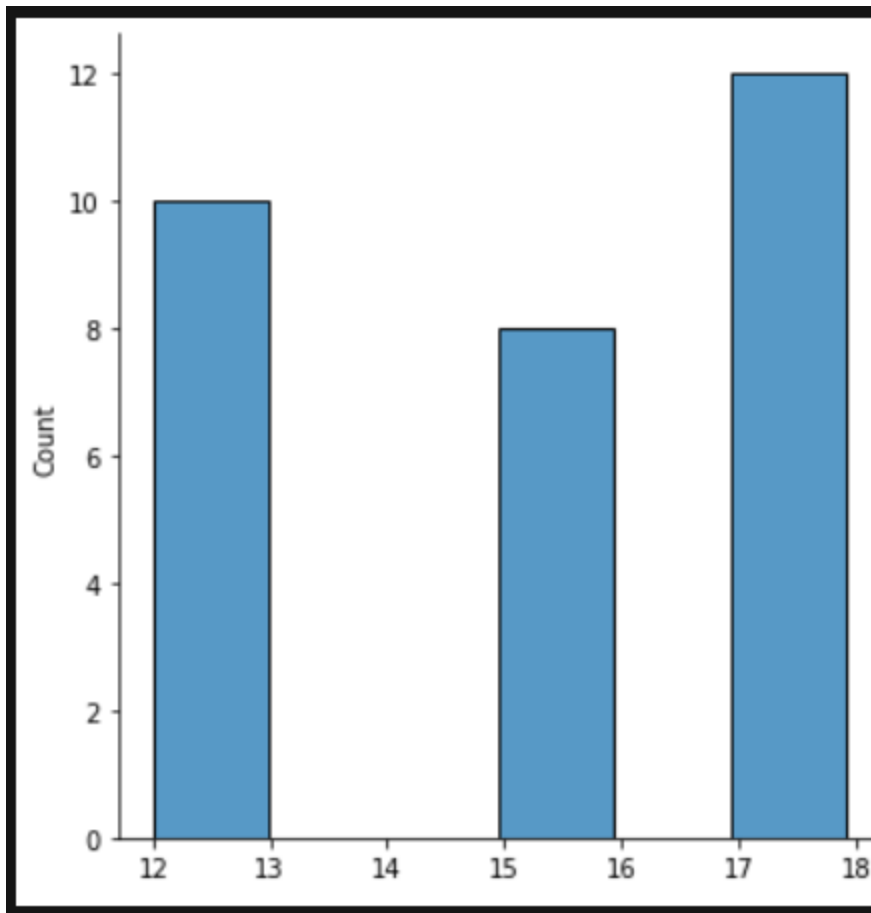
2) Used linear discriminant function to calculate the accuracy on the classification task with 80% training and 20% testing data.

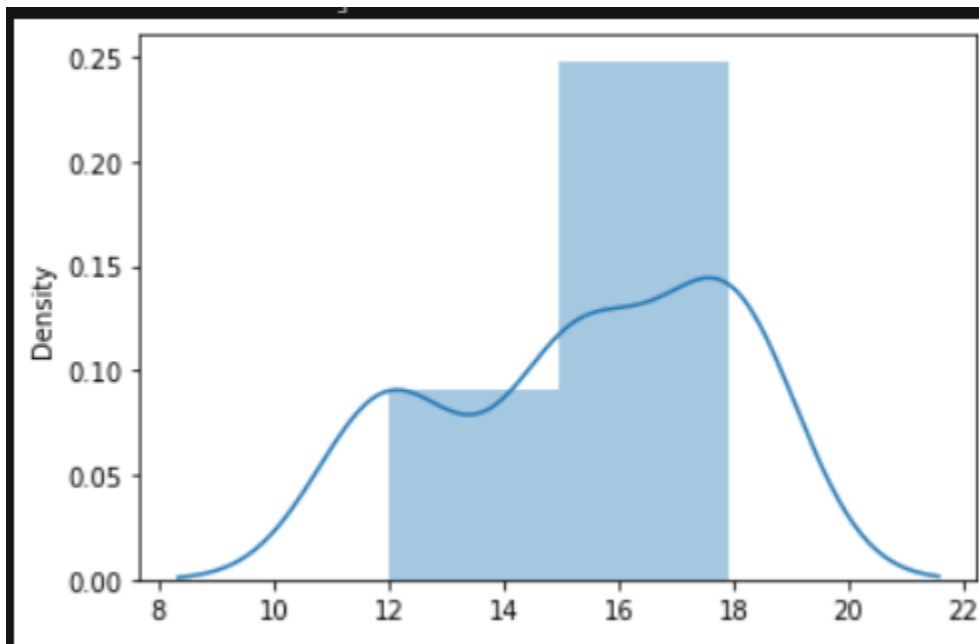
- Accuracy of 1 was achieved.

3) Bayes risk

- Bayes risk was calculated using the formula:

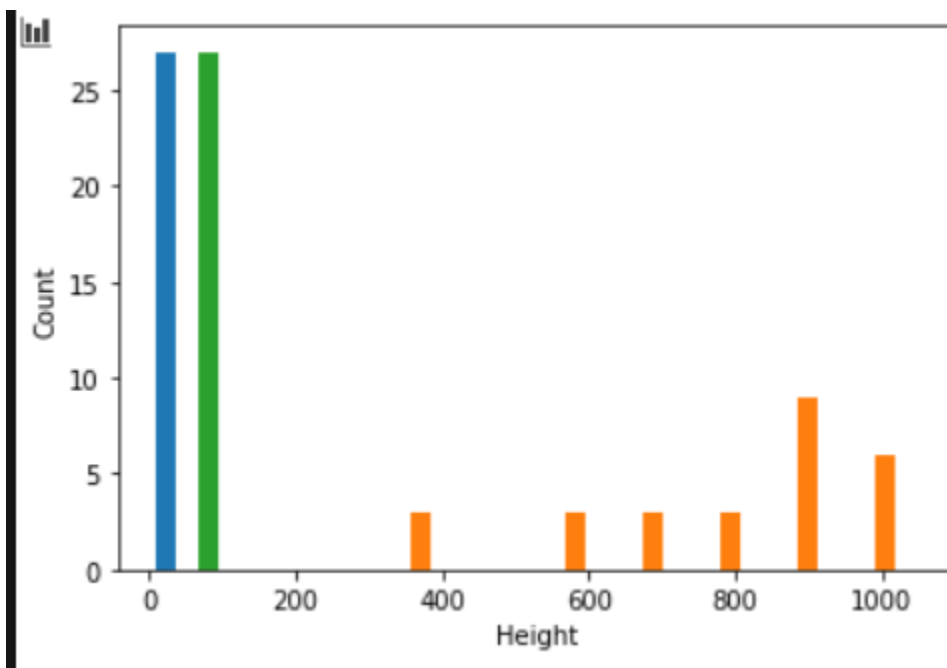
$R(\alpha_1 / x) = \lambda(\alpha_1 / w_1) P(w_1 / x)$  and summing over all the features





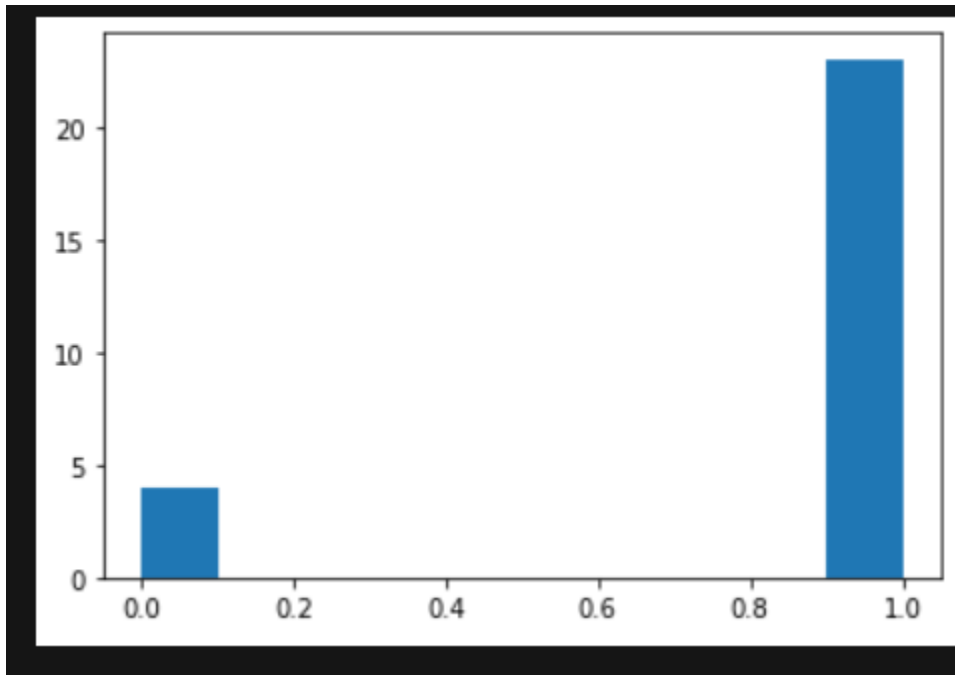
Q3)

1. Distribution of features were plotted:



- 2) Divided the model into class 1 and 2 according to their price(>550)





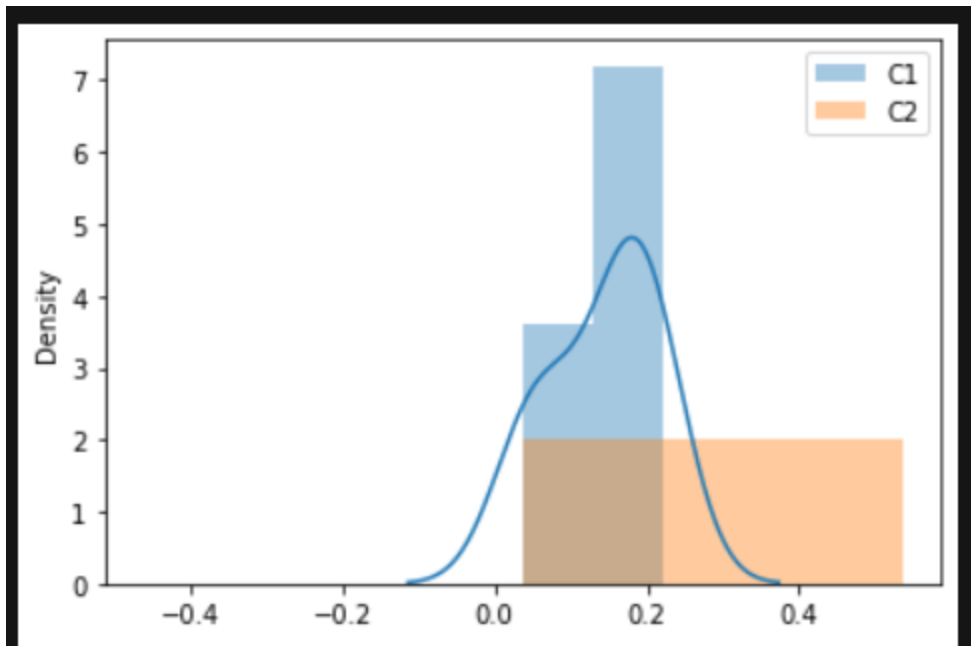
3) Prior Probabilities were calculated :

The probabilities obtained were 0.85 and 0.14 for class 1 and 2 respectively

4)Posterior probabilities were calculated for

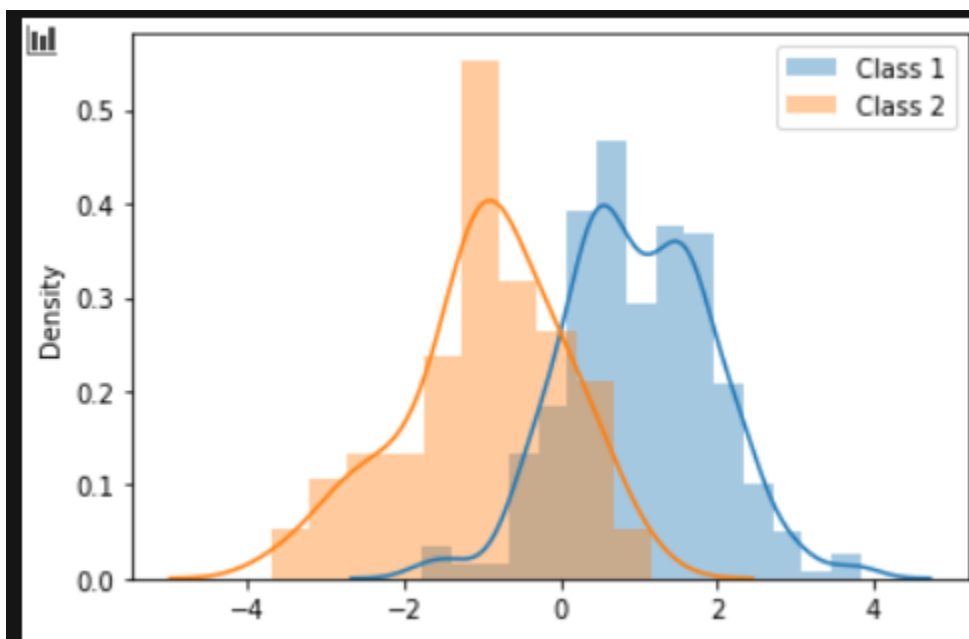
Posterior =Likelihood \* Prior/evidence

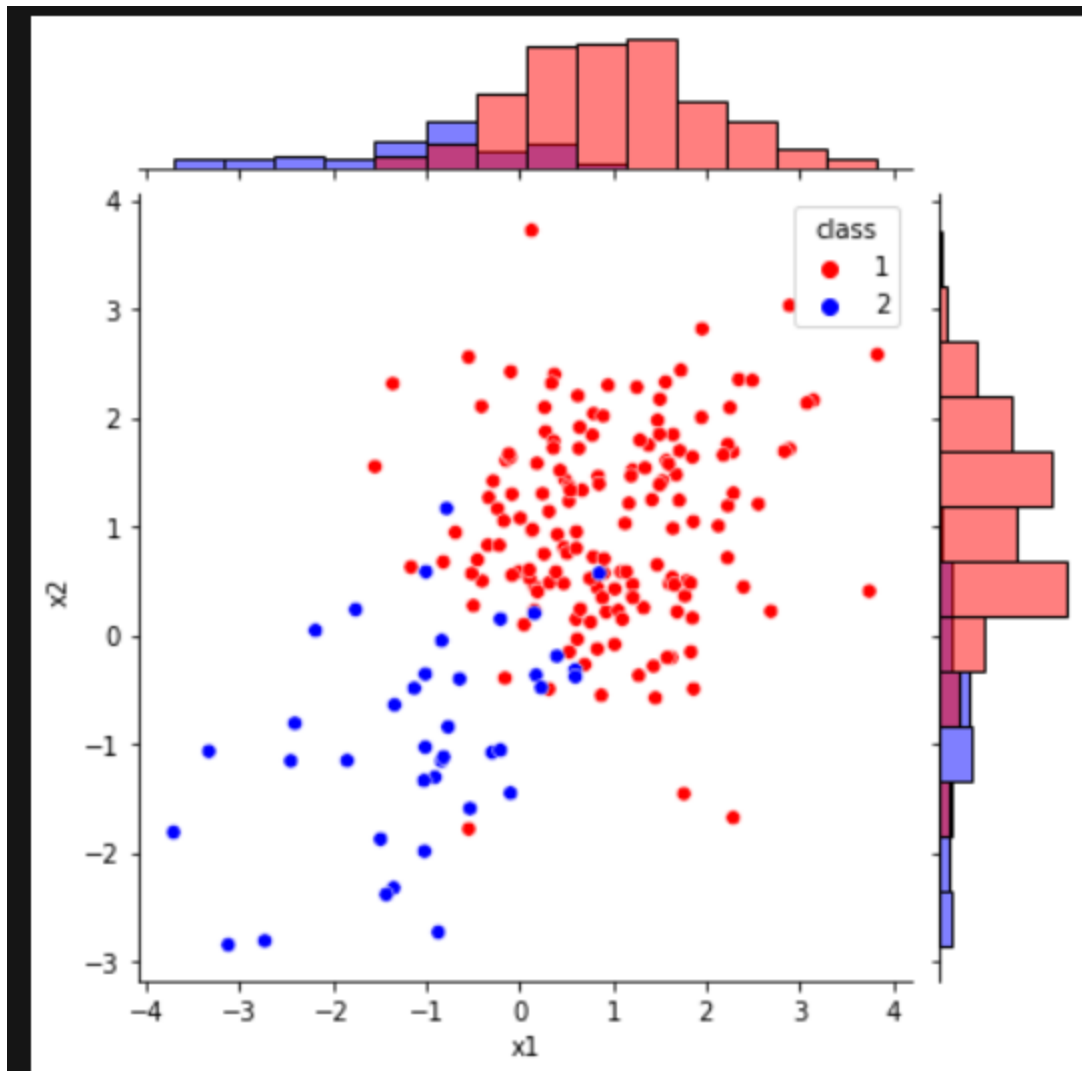
Since evidence is same the distribution obtained was:



Q4)

1. Distributions were plotted

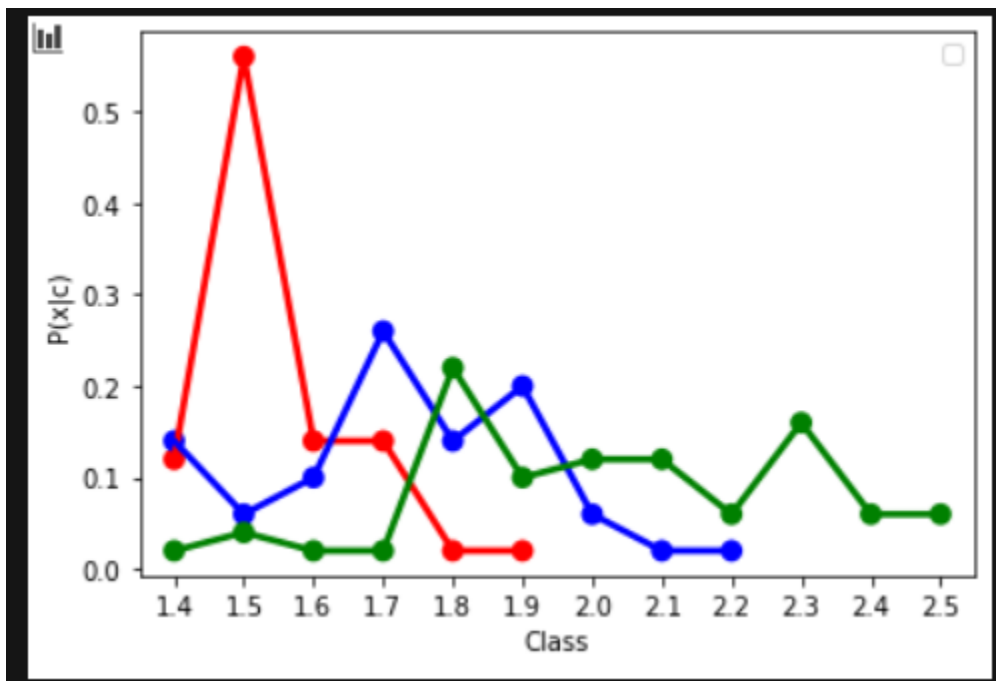
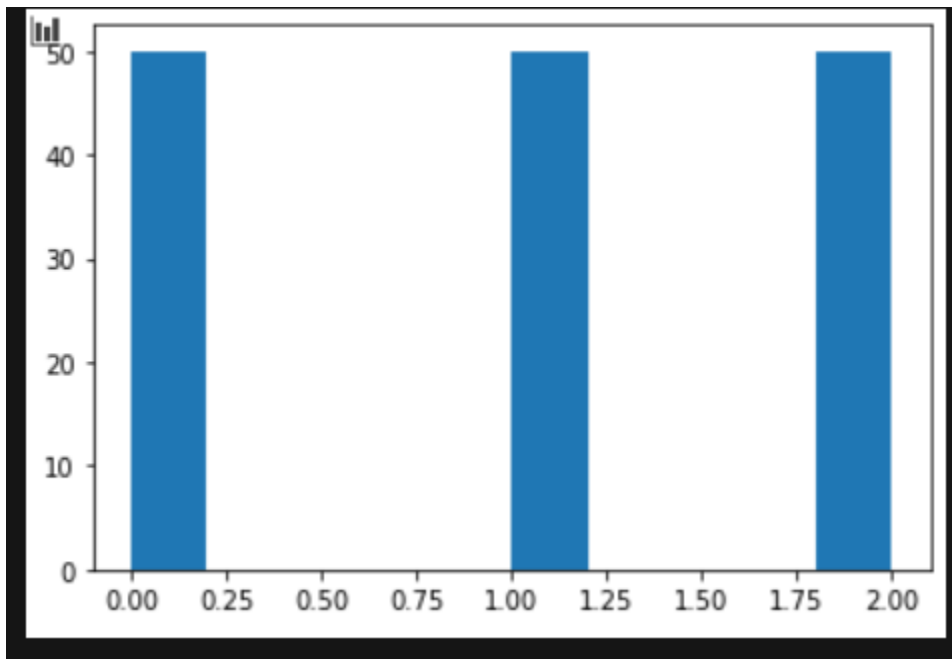


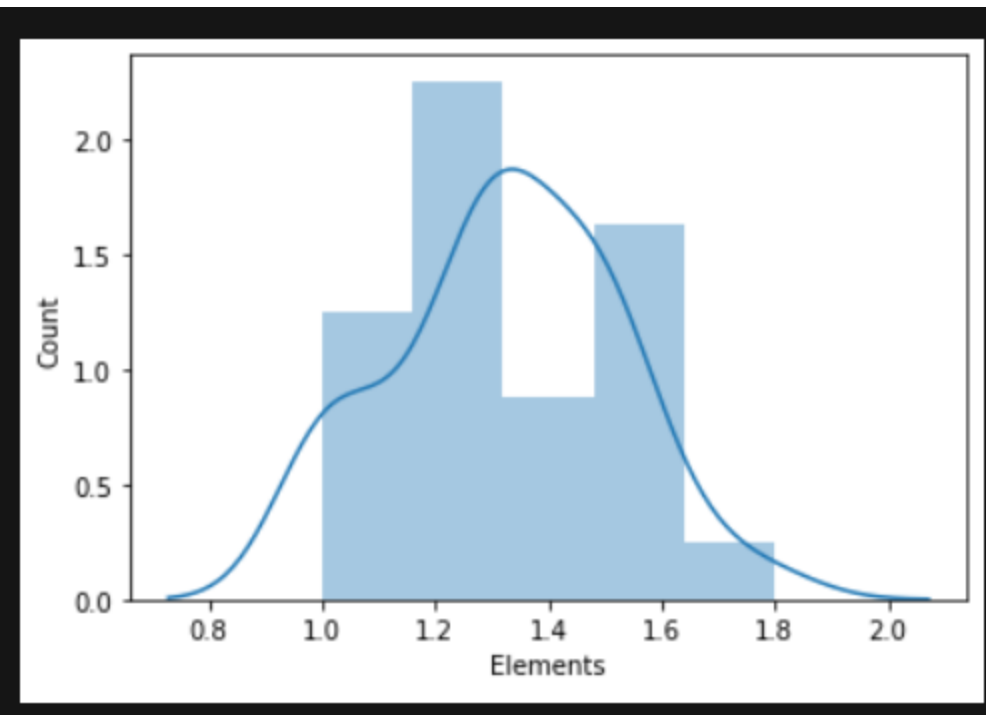
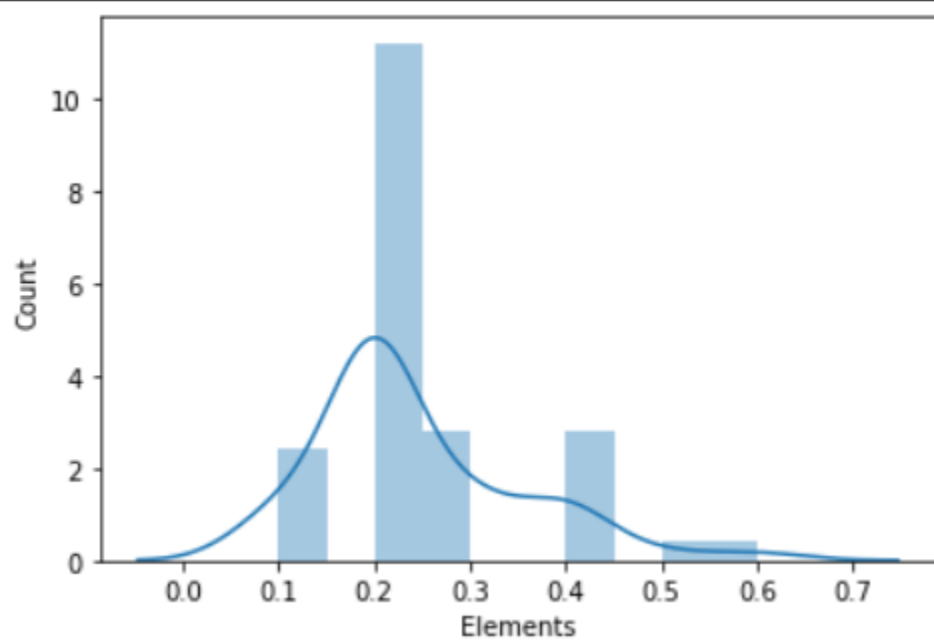


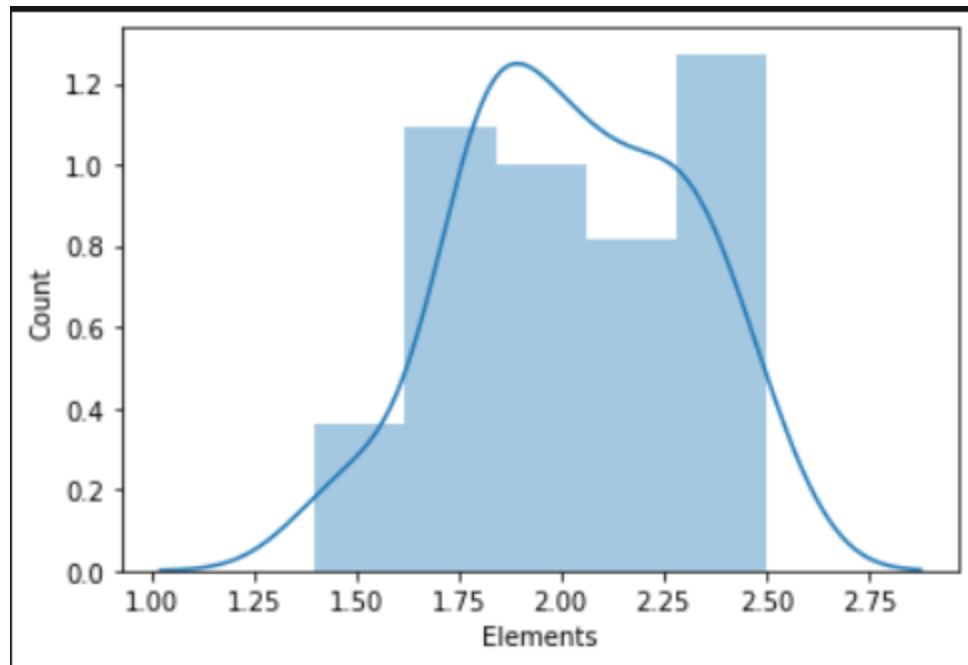
2) Posterior probabilities were calculated using the same formula.

## Q5) Real life dataset analysis

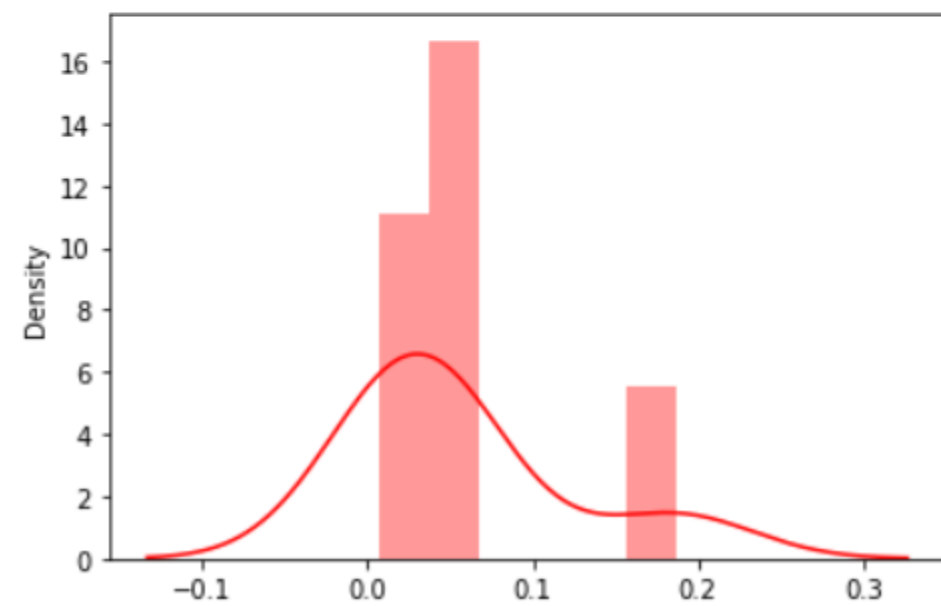
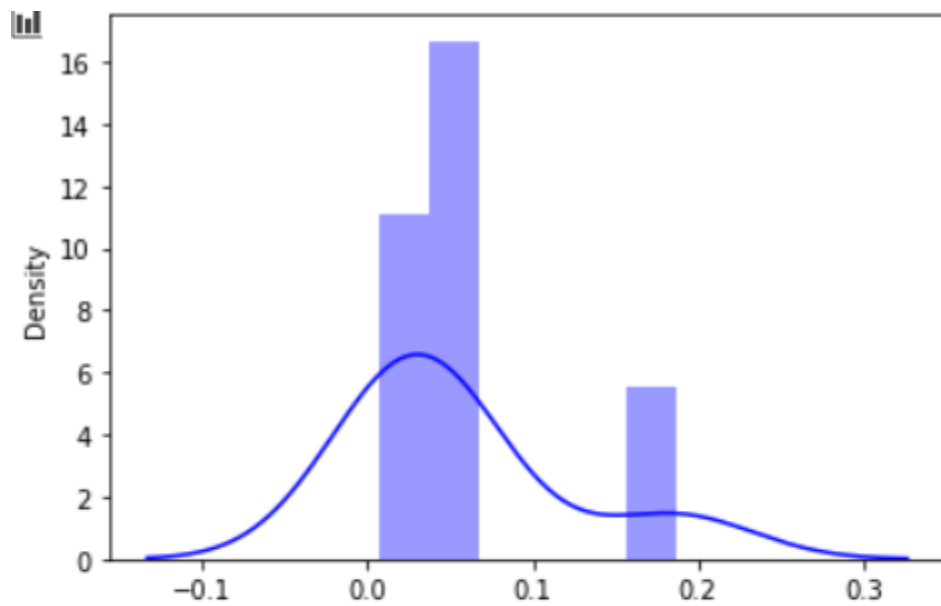
Feature - PetalWidthCm

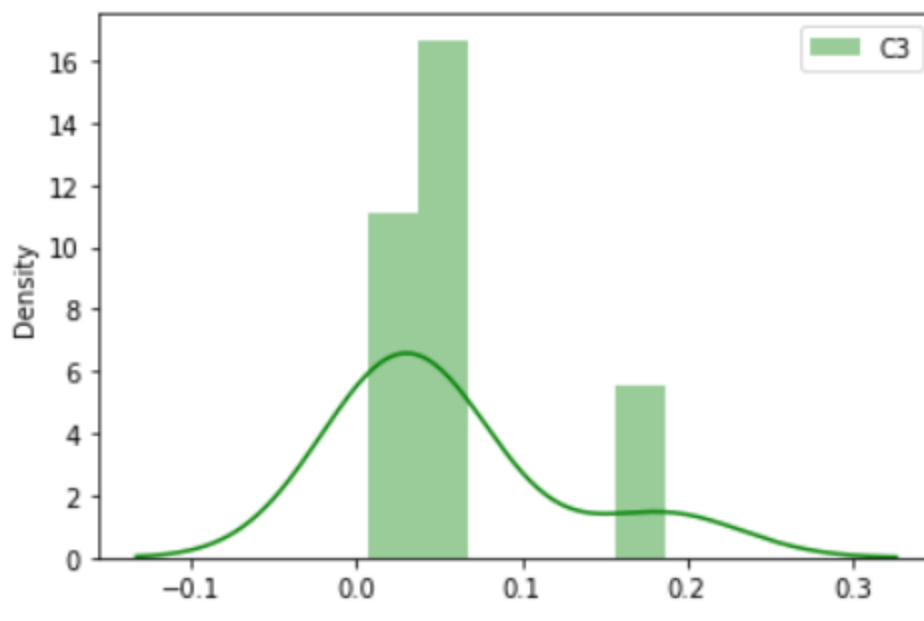






Posteriors were calculated:





1.