**PROGRAMMING ASSIGNMENT – 1**

Name:- Gowtham Kumar Kanchi

Id:- 1002044003

**NAVIE BAYES FROM SCRATCH**

**Navie Bayes: -**

The Naive Bayes algorithm is used for binary and multi-class classification problems. It assumes independence between features within a class. It's based on Bayes' Theorem, which calculates the probability of a hypothesis given evidence.

Bayes-Theorem: P(H|E) = P(E|H) \* P(H) / P(E)

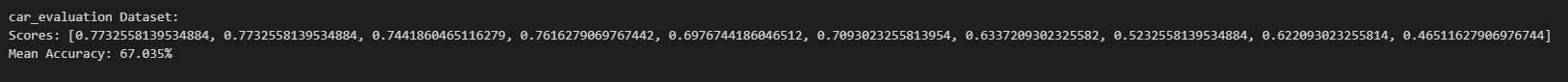
**Code Structure: -**

This code implements the Naive Bayes algorithm for classification using the car evaluation dataset. The following is a summary of the functions involved in the code:

1. **load\_data(filename)**: This function loads the data from a .csv file specified by the **filename** argument.
2. **convert\_column\_to\_float(data\_list, col\_idx)**: This function converts the values of a specified column in the dataset to floating-point numbers. The column index is specified by the **col\_idx** argument.
3. **convert\_column\_to\_integer(data, column\_index)**: This function converts the values of a specified column in the dataset to integers by mapping the unique values in the column to integer values.
4. **convert\_col\_to\_int\_car(dataset, col\_idx)**: This function converts the values of a specified column in the dataset to integers. It is specifically used for the car evaluation dataset.
5. **split\_data\_into\_folds(data, num\_folds)**: This function splits the dataset into **num\_folds** folds, where each fold will be used for testing the algorithm.
6. **calculate\_accuracy(true\_labels, predicted\_labels)**: This function calculates the accuracy of the algorithm by comparing the true class labels to the predicted class labels.
7. **evaluate\_algorithm(data, algorithm, num\_folds)**: This function evaluates the algorithm by using **num\_folds**-fold cross-validation. The algorithm is specified by the **algorithm** argument and the data is specified by the **data** argument.
8. **separate\_by\_class(dataset)**: This function separates the dataset by class values and returns a dictionary of the separated data.
9. **calculate\_mean(numbers)**: This function calculates the mean of a list of numbers.
10. **calculate\_std\_dev(numbers)**: This function calculates the standard deviation of a list of numbers.
11. **summarize\_data(dataset)**: This function summarizes the data by calculating the mean, standard deviation, and count for each column in the dataset.
12. **summarize\_by\_class(dataset)**: This function summarizes the data by class values. It calculates the mean, standard deviation, and count for each column in the dataset, separated by class values.
13. **calculate\_prob\_distribution(x, mean, std\_dev)**: This function calculates the Gaussian probability distribution function for a given value **x** and its mean and standard deviation.
14. **calculate\_class\_probabilities(class\_summary, row)**: This function calculates the class probabilities for a given instance specified by the **row** argument, based on the **class\_summary** of the training data.
15. **predict\_class(summaries, row)**: This function predicts the class label for a given instance specified by the **row** argument, based on the **summaries** of the training data.
16. **naive\_bayes(train\_set, test\_set)**: This function implements the Naive Bayes algorithm by using the **train\_set** data to train the model and the **test\_set** data to make predictions.

Finally, the code prints the scores and mean accuracy of the algorithm. The mean accuracy is calculated by taking the average of the accuracy scores obtained from each fold.

**Naive Bayes on car evaluation dataset: -**

**Python:**

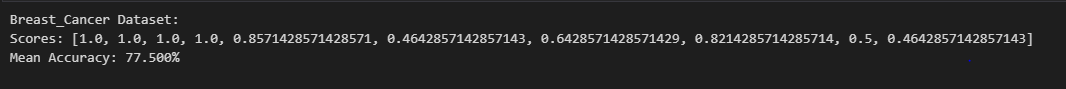
**Weka:**

Graphical user interface

Description automatically generated

**Naive Bayes on Breast Cancer dataset: -**

**Python:**

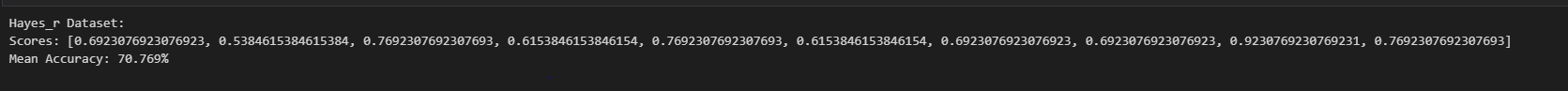


**Weka:**

Graphical user interface, text

Description automatically generated

**Naive Bayes on Hayes r dataset: -**

**Python:**

**Weka:**

A picture containing graphical user interface

Description automatically generated

**References:**

<https://www.geeksforgeeks.org/ml-naive-bayes-scratch-implementation-using-python/>

<https://machinelearningmastery.com/k-fold-cross-validation/>

<https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/>