**Introduction**

The aim of the experiment is to import the iris dataset, plot and study the dataset. Using Naïve Bayes classification technique, we are going to predict and find the confusion and accuracy metrices.

Naive Bayes classifiers are a collection of classification algorithms based on Bayes’ Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other

Bayes’ Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes’ theorem is stated mathematically as the following equation:



**Dataset**

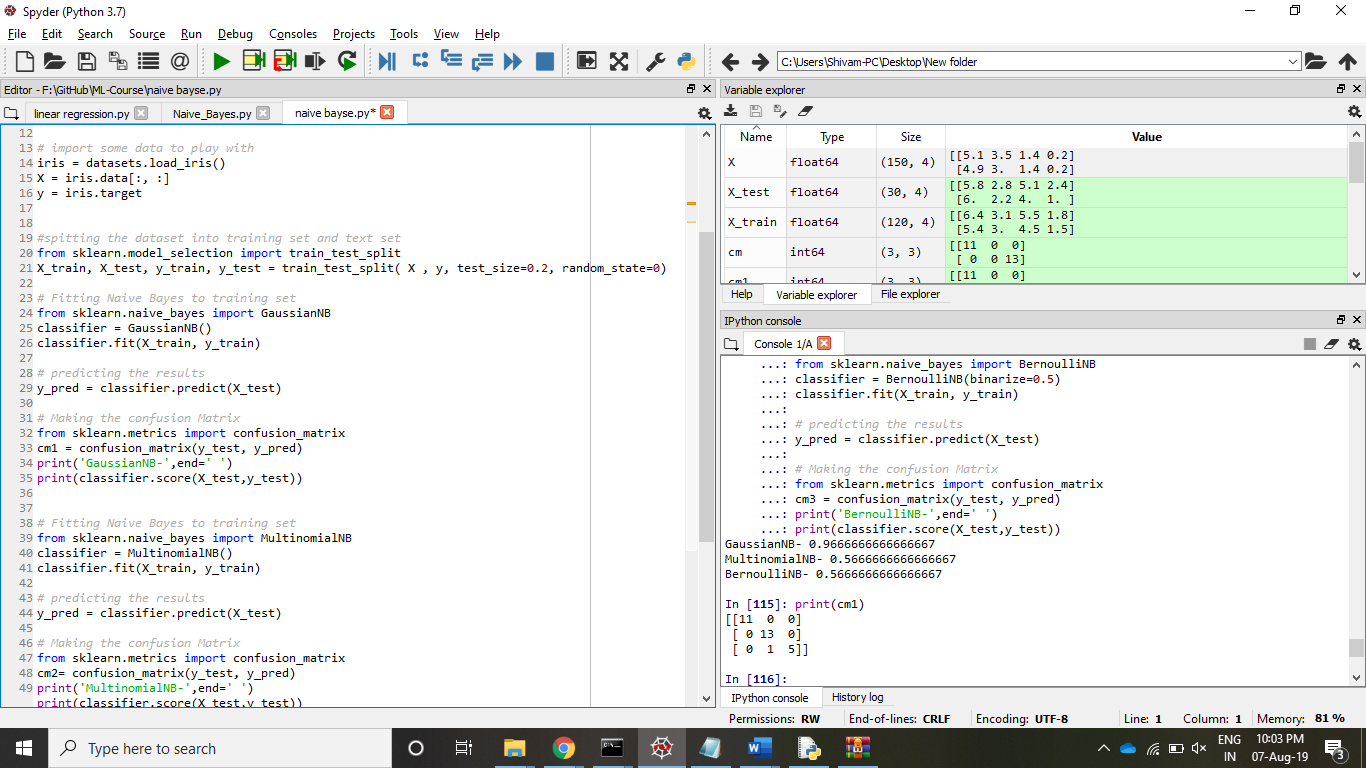
The dataset used in the experiment is the iris dataset. This data sets consists of 3 different types of irises’ (Setosa, Versicolour, and Virginica) petal and sepal length, stored in a 150x4 numpy.ndarray. The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width.

**Method**

First, we import the iris dataset. We use all the four columns -Sepal Length, Sepal Width, Petal Length and Petal Width because all the four columns effects and have weight on the result. Then using sklearn model selection module we split the data into training and testing data. The testing data comprising of 20% of dataset and training rest of 80% of dataset. Now we train the classifier using training data. And then we predict and calculate score on testing data. Gaussian naïve Bayes gives us the best accuracy. We will find the confusion matrix on the predicted and testing labels.

**Result & Interpretation (in terms of Confusion matrix and accuracy metrics)**

Using Gaussian Naïve Bayes, we get an accuracy of 96.66% in the dataset if we include all the four columns in the training data. We also found that we get the highest accuracy when we include all four columns in the training data. So all four columns are important for better predictions The snapshot of confusion matrix and accuracy metrics of all Naïve Bayes algorithms is attached below.



**Conclusion & Scope**

Naive Bayes algorithms are mostly used in sentiment analysis, spam filtering, recommendation systems etc. They are fast and easy to implement but their biggest disadvantage is that the requirement of predictors to be independent. In most of the real-life cases, the predictors are dependent, this hinders the performance of the classifier.