**Python and Deep Learning**

**CSEE5590/490 APS**

**Lab Assignment -2**

**Submitted by:**

**K.N.V.Saiindra Kumar**

**T.V.S.Sai Ram**

1. **Introduction**

Searching a dataset and plotting categories available in the dataset.

Programming a code to implement support vector machine using svc with rbf kernel.

Reading the input file and applying some functions like lemmatization and bigrams and calculating word frequency and repeated bigrams in the file.

Noticing the changes in the k nearest neighbour algorithm by changing the k values.

1. **Objectives**
   1. We tend to apply packages like matplotlib and load iris dataset and data science libraries like scikit learn package.
   2. We focus on checking and using gamma function and importing svc function.
   3. Reading the file using input function and using concepts like lemmatization and bigrams and word frequency and finally summarization of the file is done.
   4. We load the dataset and using kneighbour classifier in that library and. the range is determined 2 times in order to notice the changes
2. **Approaches**

We used many libraries like matplotlib, sklearn, metrics, nltk, bigrams, svm, GaussianNB to predict the accuracy for the k values, and to visualize the dataset , nltk to call lemmatizer, stemmer, and bigram functions.

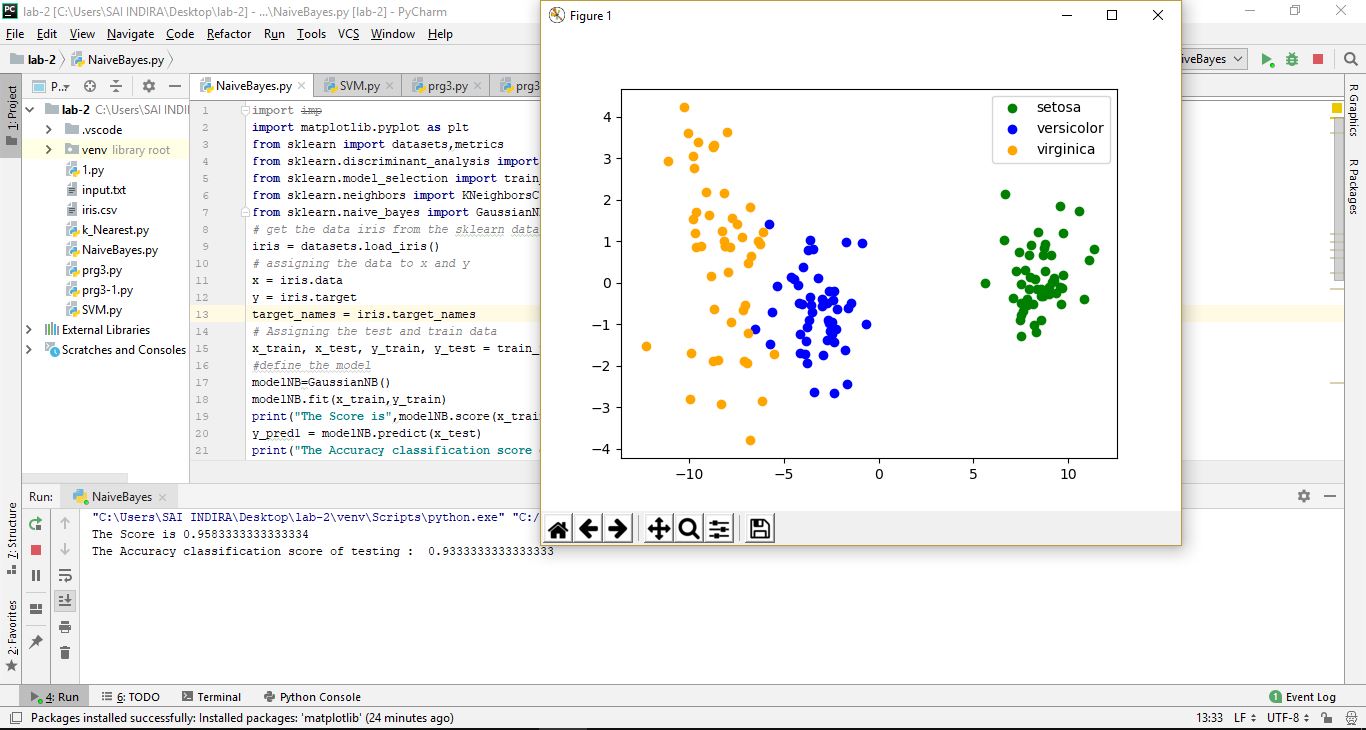
1. **Workflow**

**Question 1**

We imported all the package libraries and loaded the dateset and assigned the values to the variables, and we also assigned the training and the test data and then we called the gaussianNB function with the kneighbour classifier to predict the model and visualize them.

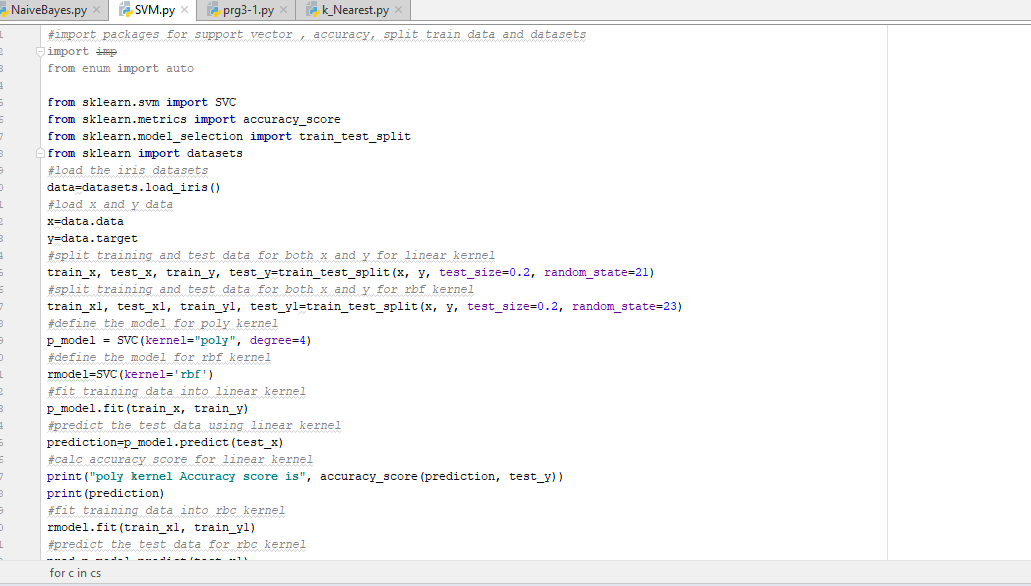


**OUTPUT**

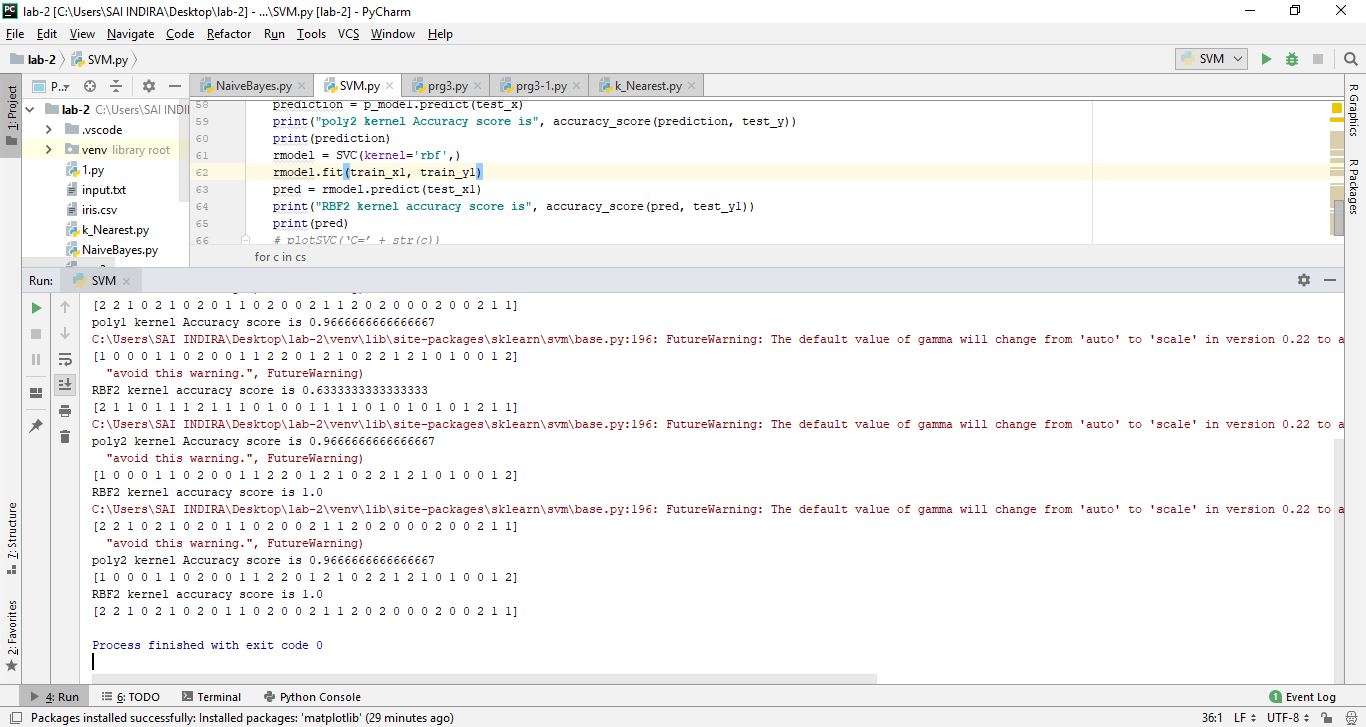


**Question 2**

We loaded the scikit learn package to call the svc functions and selection models and test and training splits and also the data sets and we called the lmodel, rmodel and also the rbf kernel and gamma value to predict the accuracy and also the if need the plot can be done.

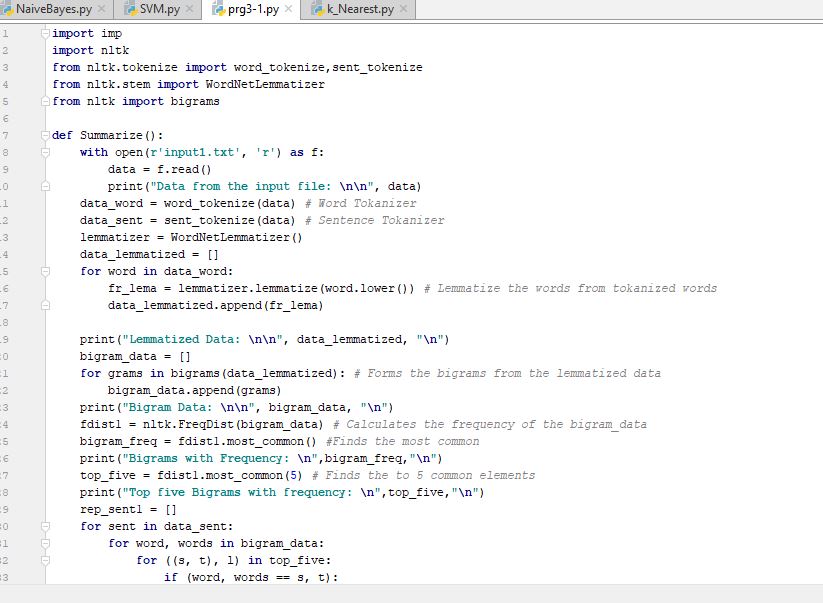


**OUTPUT**

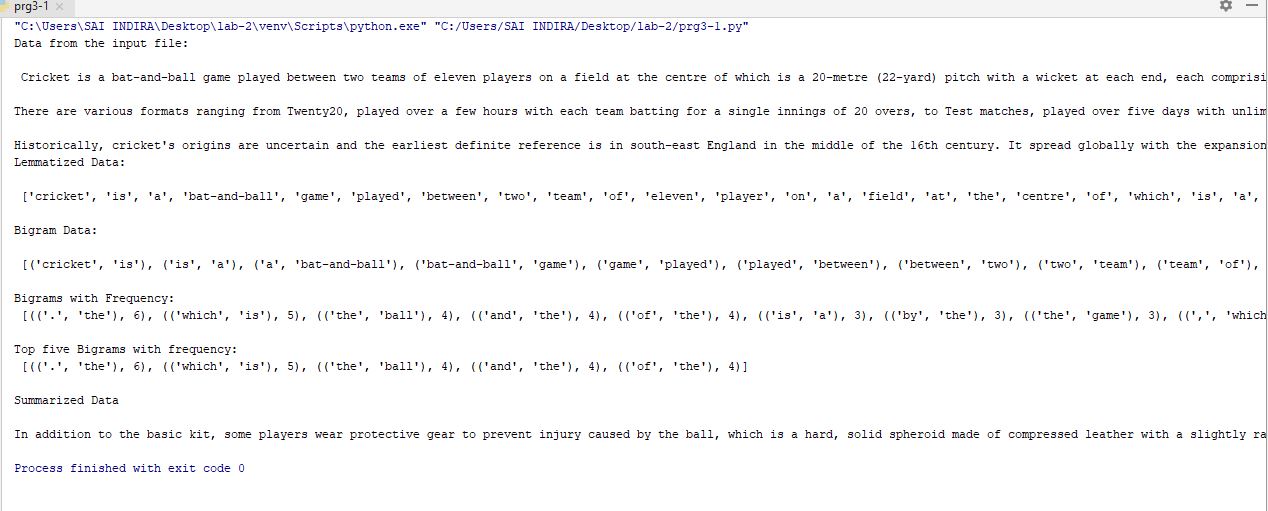


**QUESTION 3**

In this program we called the open function on the input file and the usage of nltk package to call the lemmatizer, stemmer, bigrams to determine the output on the given input file.

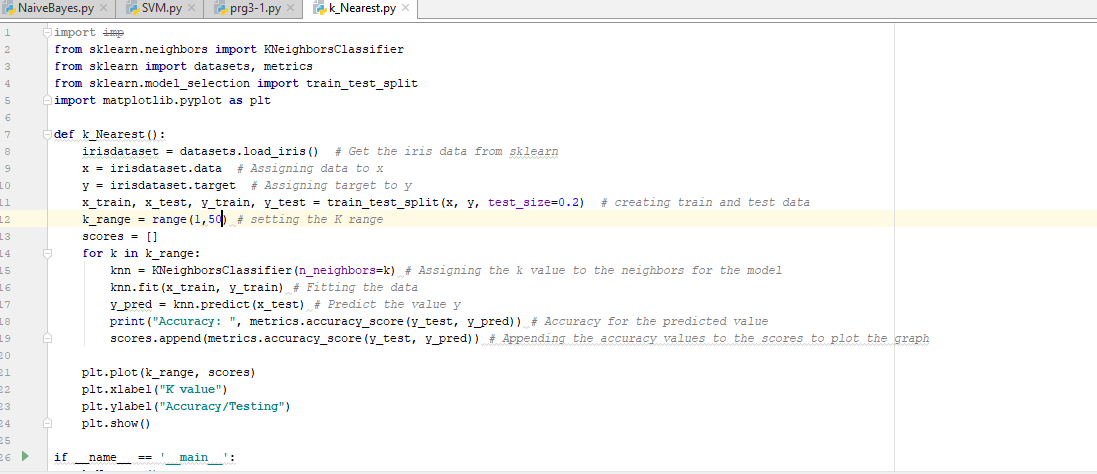


**OUTPUT**



**QUESTION 4**

In this question k nearest neighbour classifier is used on the loaded dataset and also the training and test set on the variables and the accuracy is found twice depending on the k range values.



**OUTPUT**

