**CSEE5590/490: Python and Deep Learning Programming (2018 Fall)**

**LAB ASSGINEMENT 2**

**Team ID:- 13**

**Partner 1:** Kamal Tej Veerapaneni Id - 31

**Partner 2:** Vinay Maturi Id - 17

YouTube Link: <https://youtu.be/xSwrpo3Yhlo>

**Introduction:**

In this Lab Assignment we have worked on the following tasks.

1)Pick a dataset and plot each category that is available along with creating a prediction

model and evaluating it

2)Implement Support Vector Machine Classification

3)Take an Input file and perform lemmatization, bigram and its frequency…etc

4)Reporting views on the KNN algorithm and how it changes with K

**Objectives:**

Applying dictionaries, array, and iterations in strings for finding the first non-repeated characters in string.

Looping, splitting and respective methods are used for removing contents from file\_1 which are present in also file\_2

Creating 2 lists for each class and then using a loop we can remove the members from python class list who are in common with the web development class.

Using various classes, constructors inheritance and attribute keywords like \*\*kw different sections of the task were completed.

Downloaded the html webpages which contain a table and parsed it using beautifulsoup library.

**Question 1:**

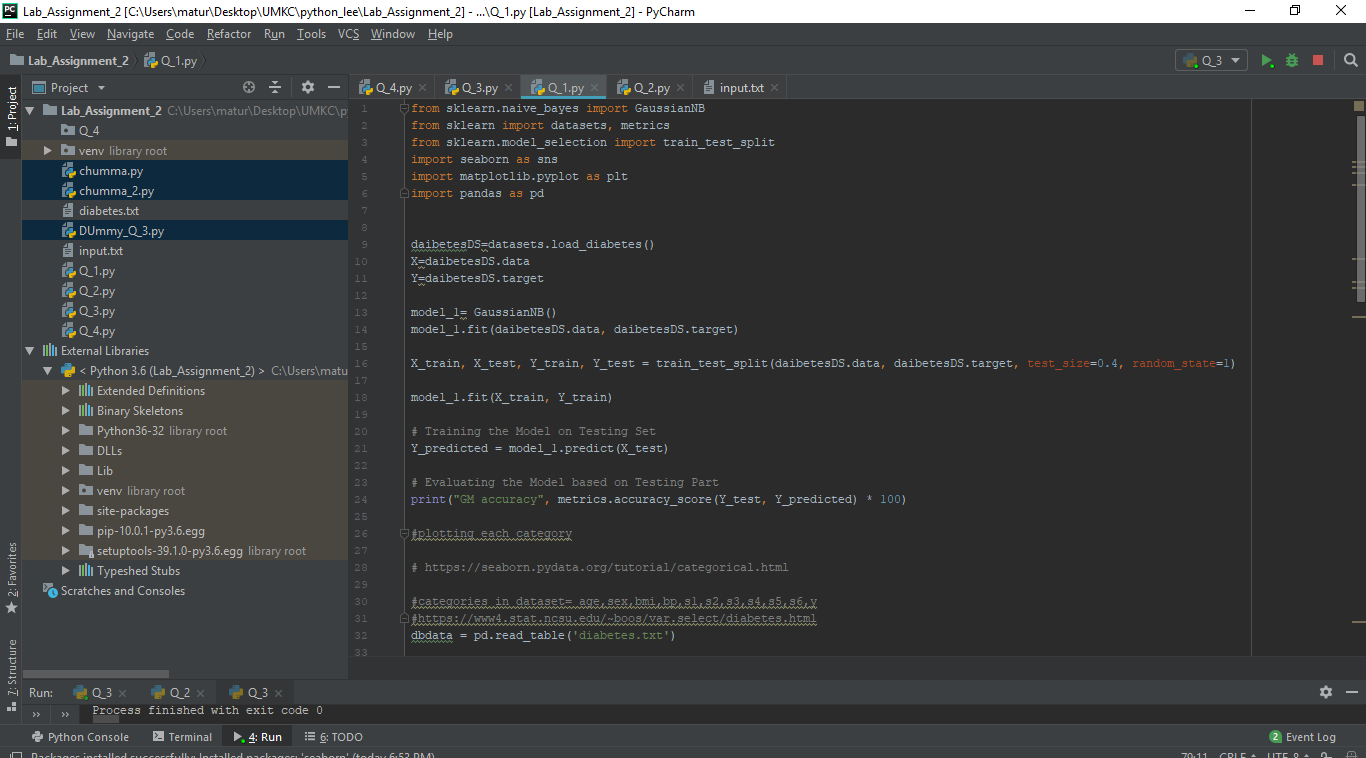
Pick any dataset from the dataset sheet in the class sheet

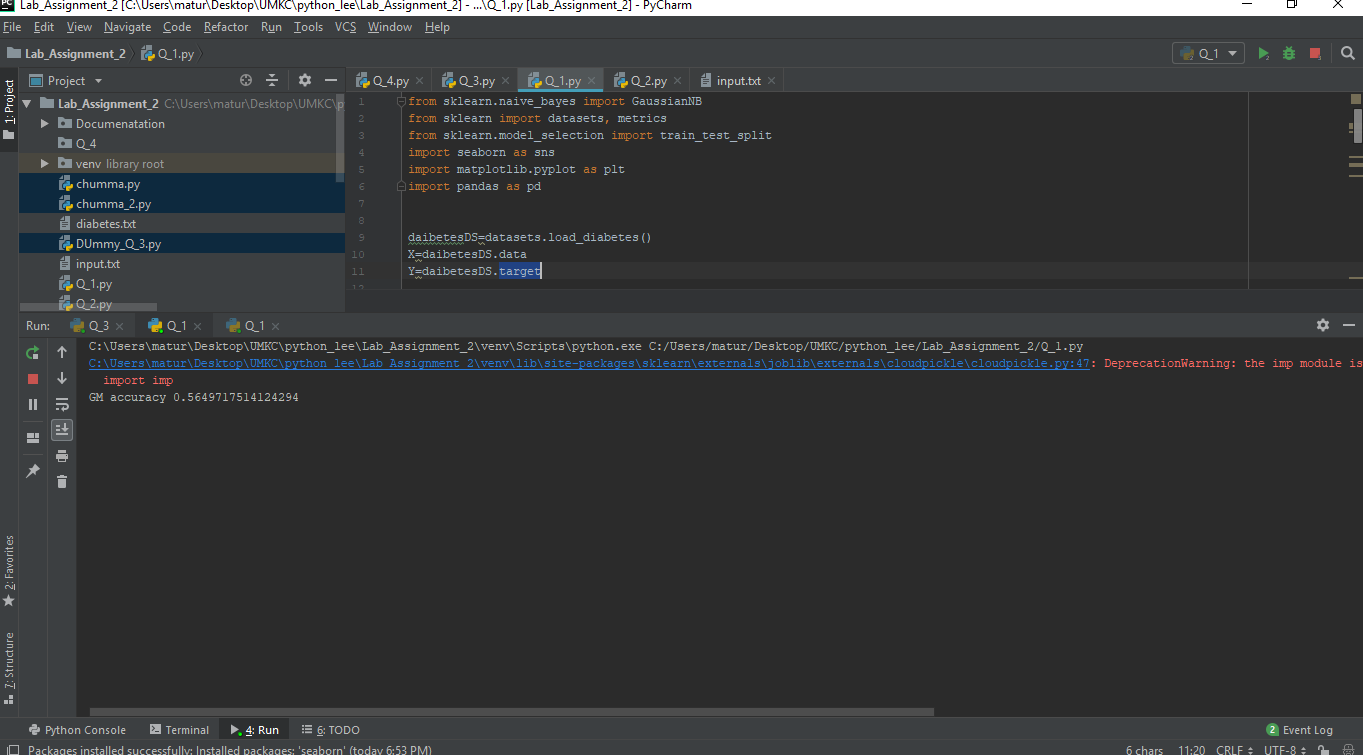
a) Plot how many of each category is available in your dataset (you can use seaborn library or matplotlib

b)Create one prediction model based on Naïve Bayes Classification and evaluate your model

**Workflow:**

The result is predicted using a NievBayers and the resultant prediction is plotted with all the catagories that are present in the dataset

**Code: Snippet:**

**Output:**

**Question 2:**

Implement Support Vector Machine classification:

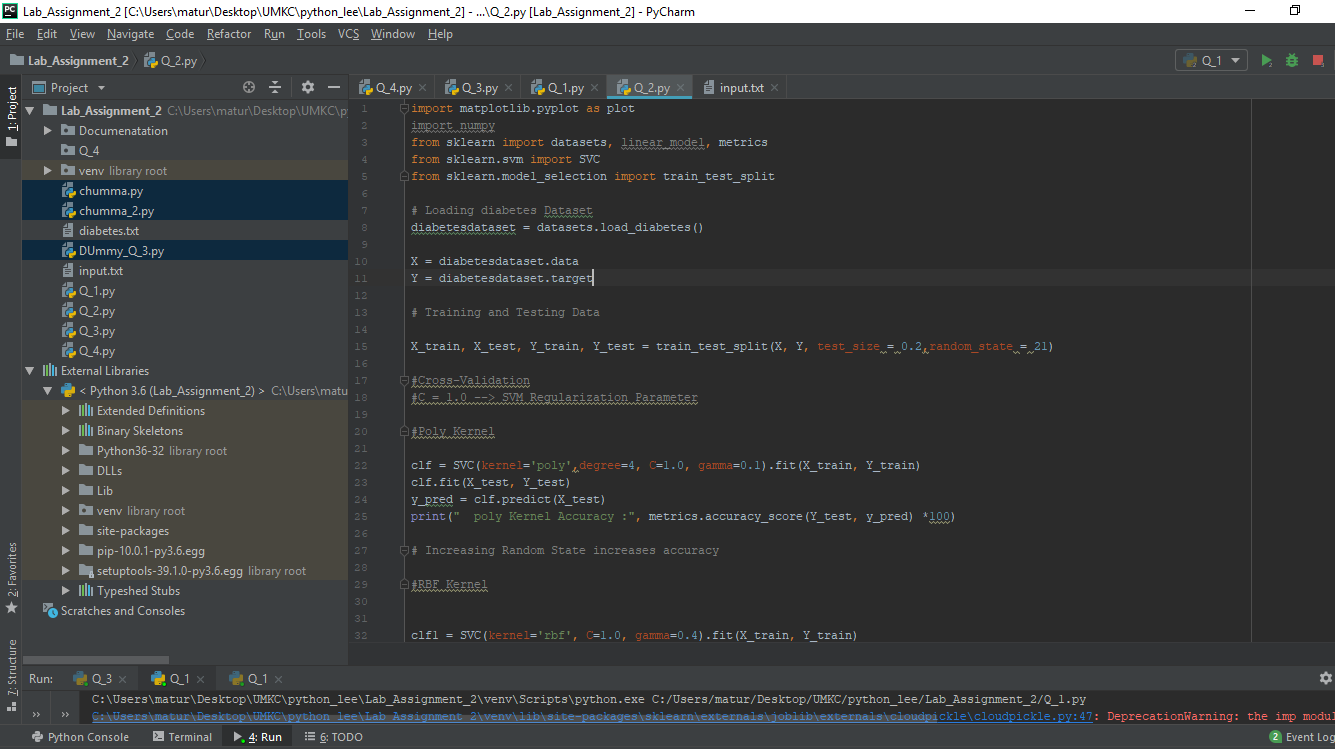
a). Apply SVC with kernel“poly” degree =4

b). Apply SVC with “rbf”kernel

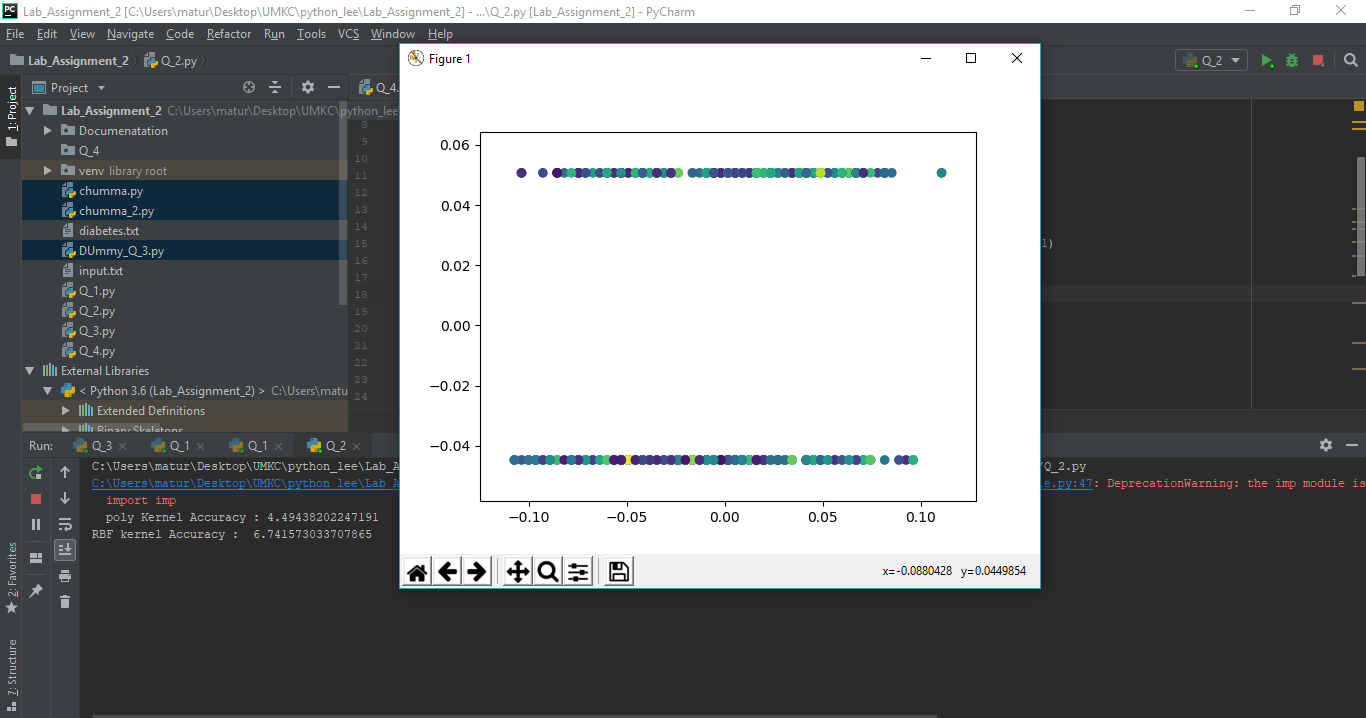
c). change gamma and C parameters in the model to see how the result may change

d). Report the accuracy of the model on both models separately andwith which parameters you got better result.

**Workflow:** SVM model is implemented on diabetes dataset by importing it and then by applying SVC with kernel “poly” and “RBF, the results are compared. Fomr the outputs we can say that the RBF model gives more accuracy than poly model.

**CODE:**

**Output:**

****

**Question 3:**

Write a program in which take an Input file. Use the simple approach below to summarize a text file:

a. Read a file

b. Apply lemmatization on the words

c. Apply the bigram on the text

d. Calculate the word frequency (bi-gram frequency) of the words (bi-grams)

f. Choose top five bi-grams that have been repeated most

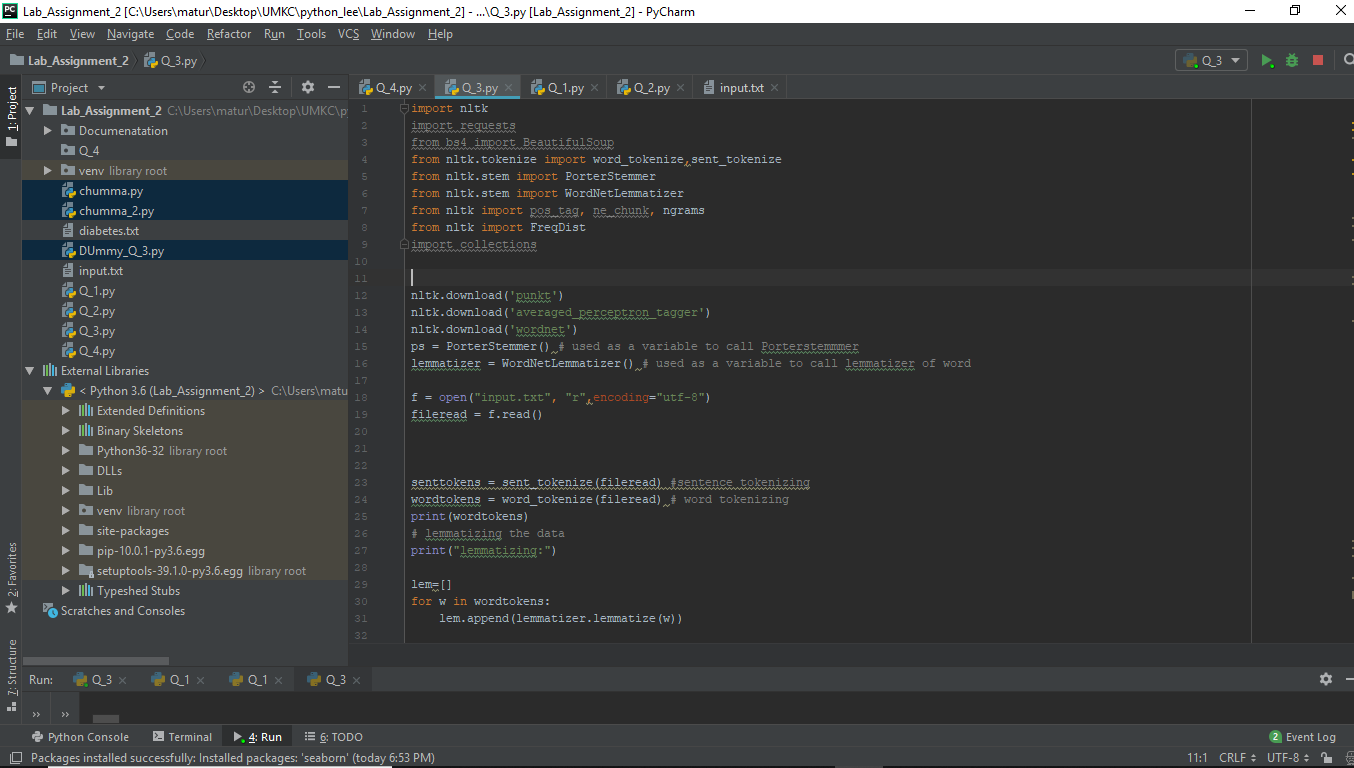
g. Go through the original text that you had in the file

h. Find all the sentences with those most repeated bi-grams

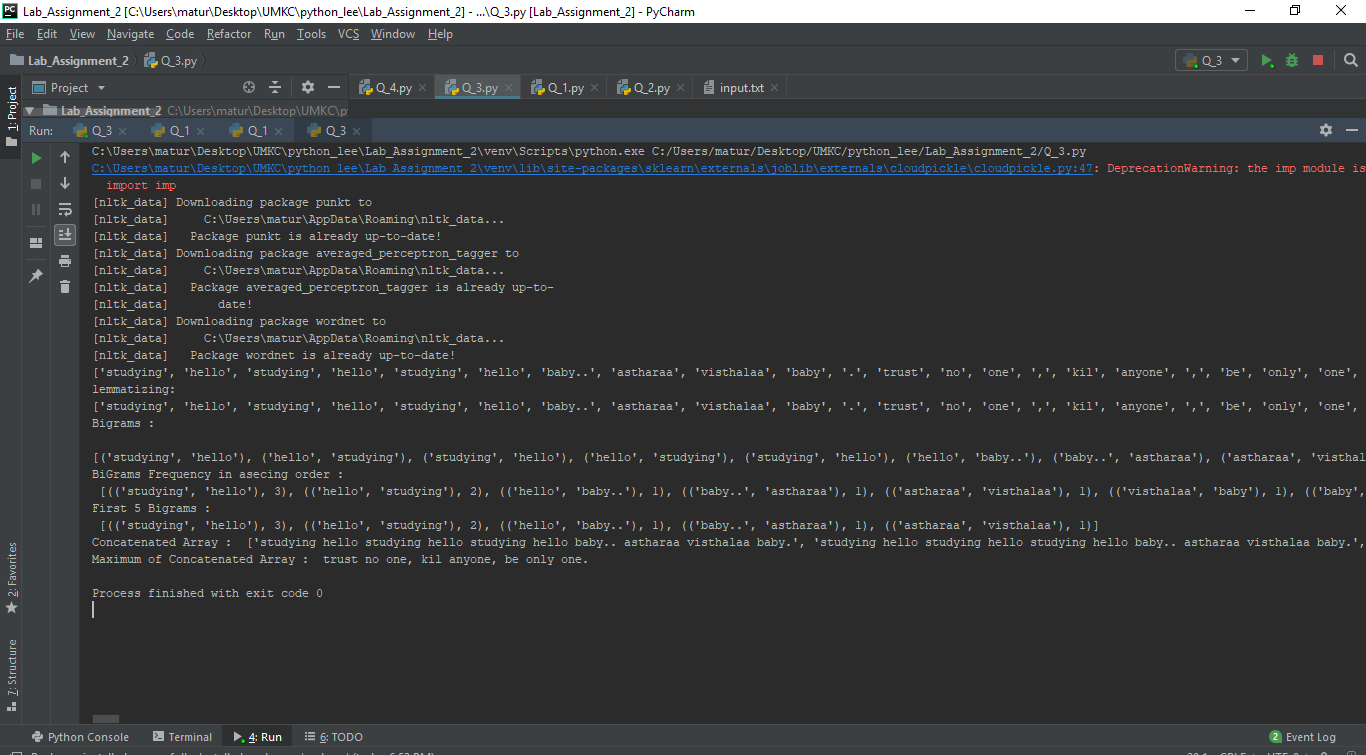
i. Extract those sentences and concatenate

j. Enjoy the summarization

**Workflow:** A file is taken as an input and limmitization is applied to the word present in the text file, then bigrams are generated from the file. The top most repeated bigrams are found and then the respective sentences containing the top bigrams are concatenated.

**CODE:**

**Output:**

****

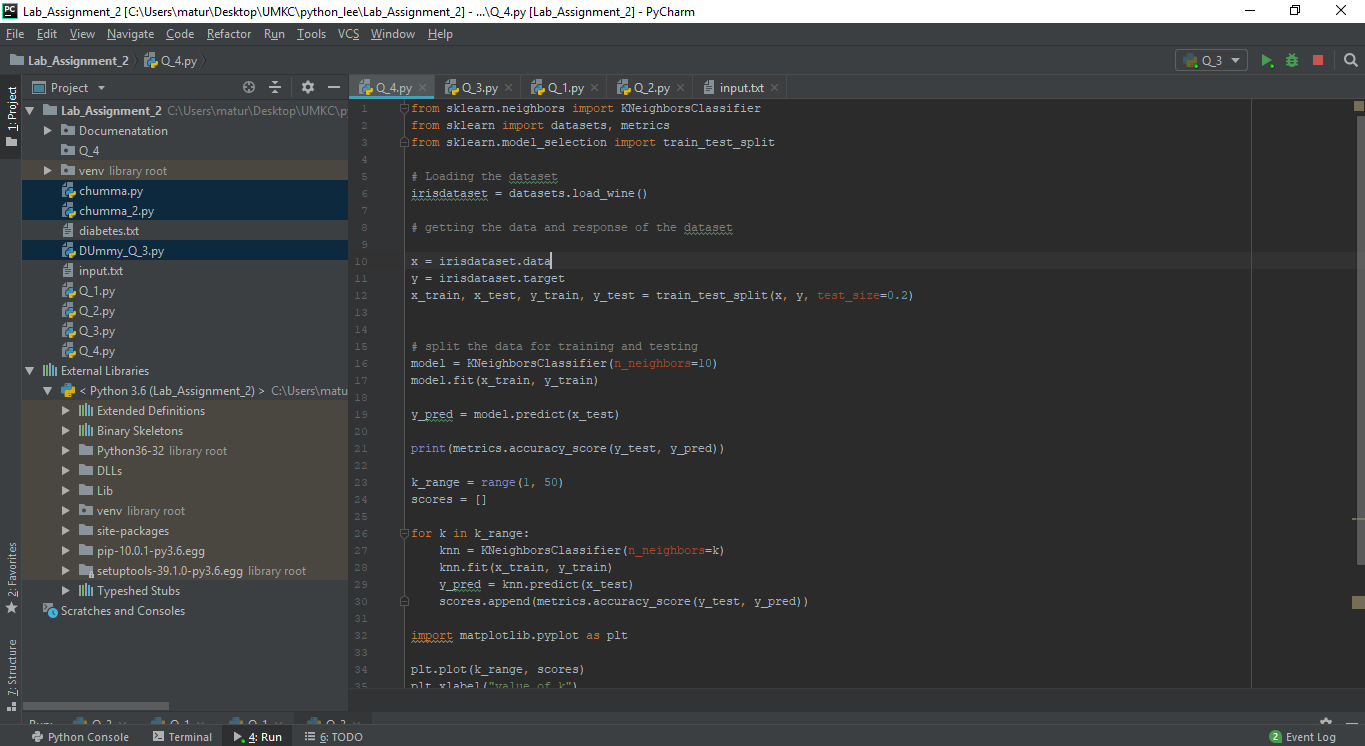
**Question 4:**

Report your views on the k nearest neighbor algorithm when we change the K how it will affect the accuracy. Provide a good justification for the changes of the accuracy when we change the amount of K. For example: compare the accuracy when K=1 and K is a big number like 50, why the accuracy will change.

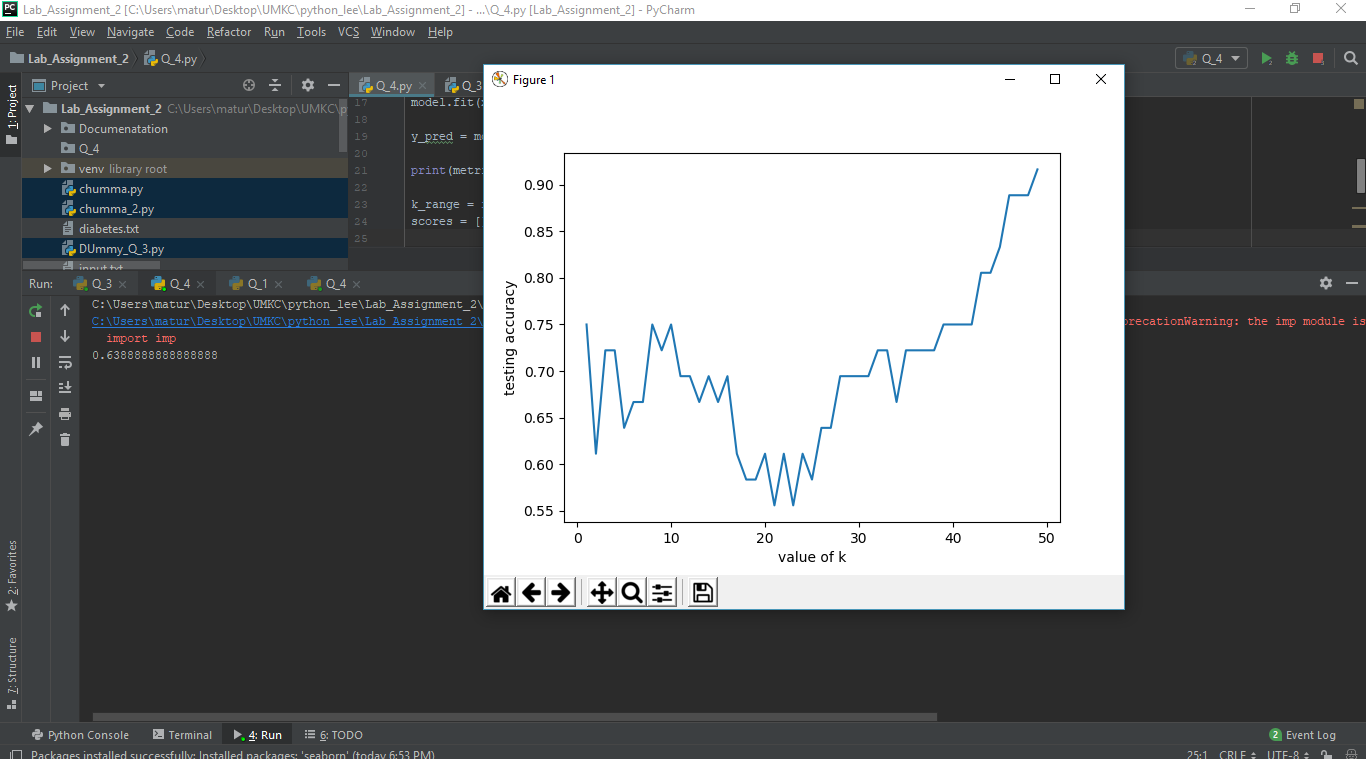
**WorkFlow:**

As the knn gives different accuracies with different k-values.The accuracies are observed as follows. The accuracies completely depends on how the model divides the dataset into trainginset and test set and k values. One of such outputs is as follows.

**Code Snippet:**

****

**Output:**

****

Thank you..