**LAB ASSIGNMENT 2**

**Name:Naga Sirisha Sunkara**

**Class ID:34**

**Team ID:4**

**Technical partners details:**

**Name:Vinay Santhosham**

**Class ID:28**

**Introduction**

We have learnt about different machine learning algorithms and usage of natural language tool kit.

**Objective**

The main objective of this assignment is to implement the machine learning algorithms and compare the different algorithms and get the accuracy scores and to implement the natural language tool kit.

**Task1:**

Comparison of Linear Discriminant Analysis and Logistic Regression classification Algorithms:

For this task we have taken the iris dataset from sklearn library. we have split the data into training and testing datasets.

Steps to make prediction using LDA:

1.Create the object for Linear Discriminant Analysis

2.Train the model based on the training dataset created.

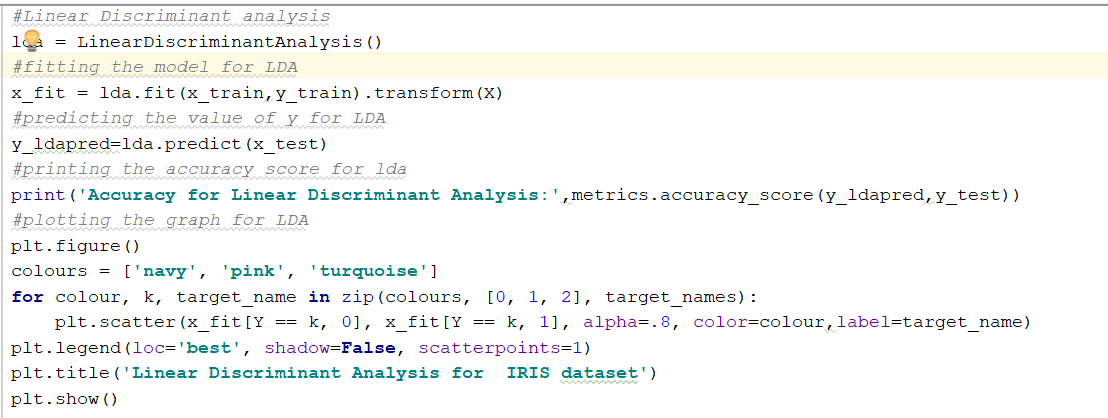
3.Predict the value with the test dataset using predict() function.

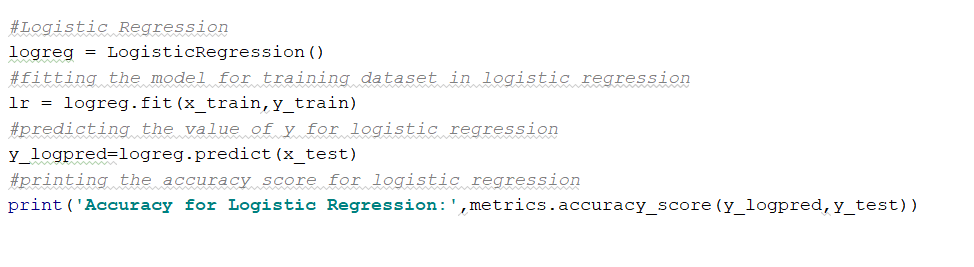
4.Find the accuracy score using metrics.accuracy\_score() function.

**Report:**

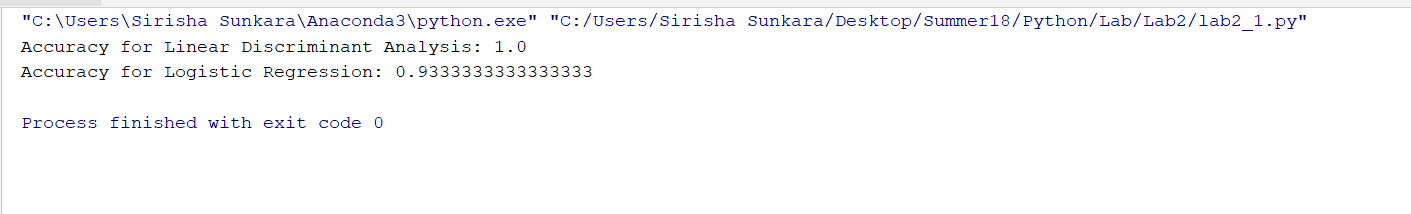
The accuracy value is more for Linear discriminant analysis in which we have taken 3 classes for classification when compared to the logistic regression.Therefore,LDA can be used when the dependent variable has 2 or more groups but logistic regression can be used only when we have two categories and to find the probabilities of these two categories.

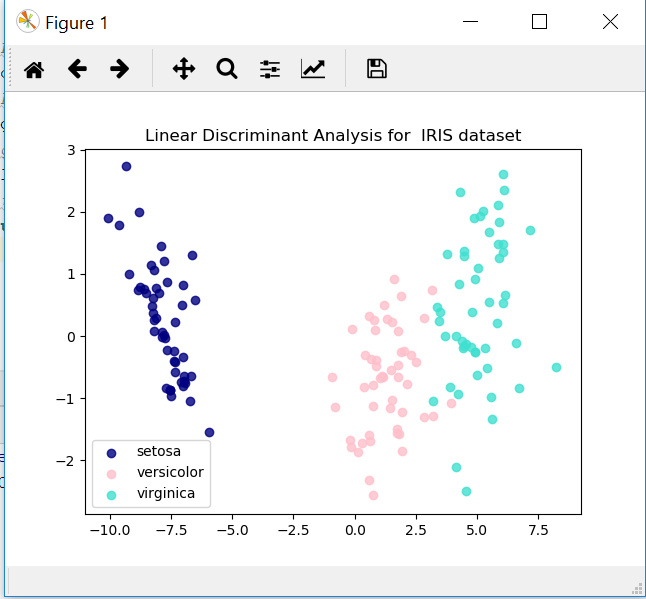
**Code:**





**Output:**

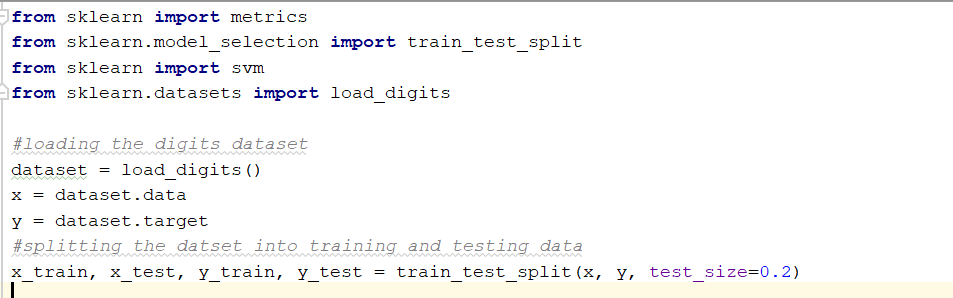




**Task 2:**

Support Vector Machine Implementation

For SVM,we have chosen the digits dataset from sklearn library.We split the data into training and testing data with 20% test data.



**Linear kernel Model**

Create the object for svm with kernel as linear and train the dataset with the linear model and predict the value and get the accuracy score.

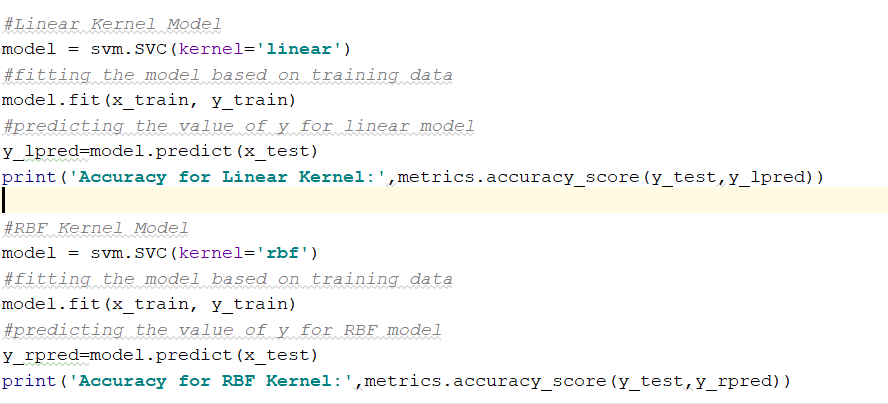
**RBF kernel model**

Create the object for svm with kernal as rbf kernel and train the dataset with the rbf kernel model.Predict the value with the trained model and get the accuracy score. Accuracy can be improved by normalizing the data and also by using the BaggingClassifier from sklearn.ensemble.

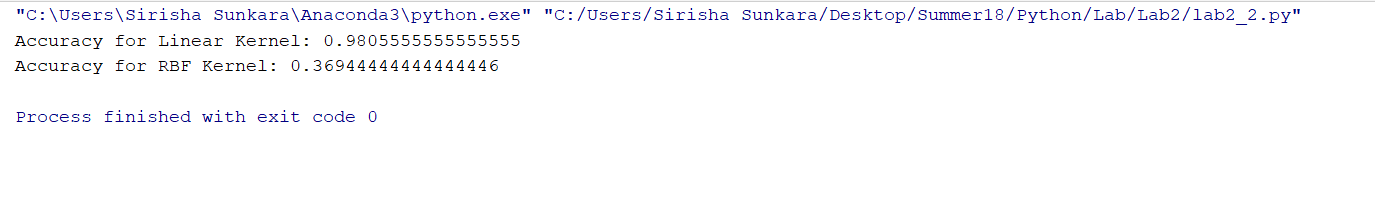
**Report**

Accuracy for svm classification with linear model is having high accuracy compared to the rbf kernel model with respect to the digits dataset.

**Code**

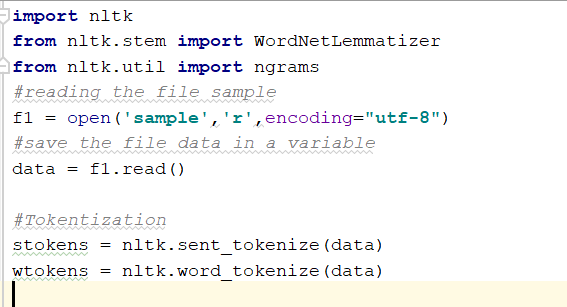


**Output**



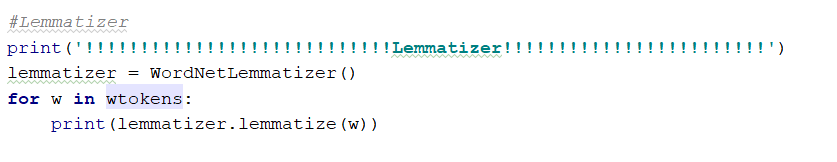
**Task3**

Usage of Natural Language Tool Kit We have taken the sample text in a file and saved the content of the file to a variable and applied word and sentence tokentization on the data.

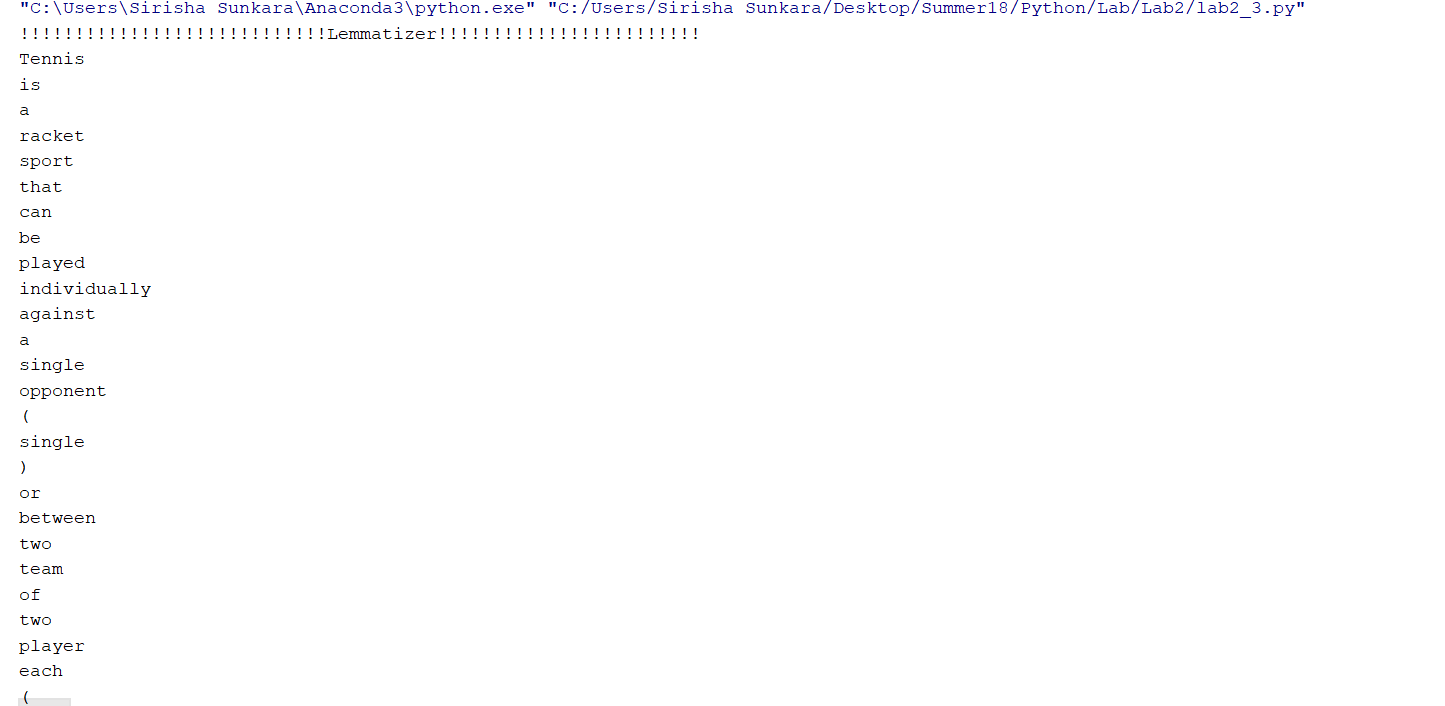


###Lemmatization

**Code**

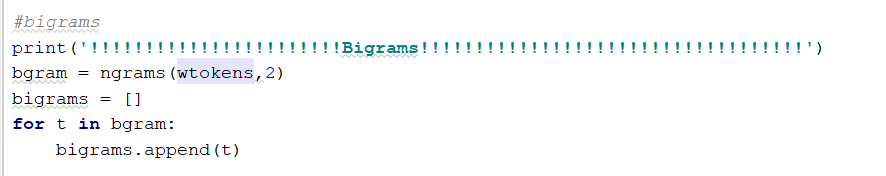


**Output**

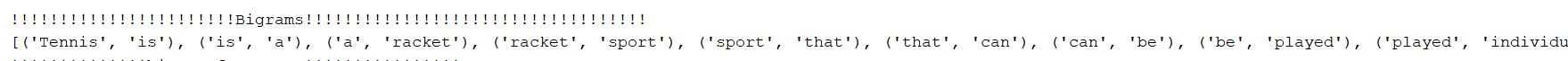


**Bigrams**

**Code**

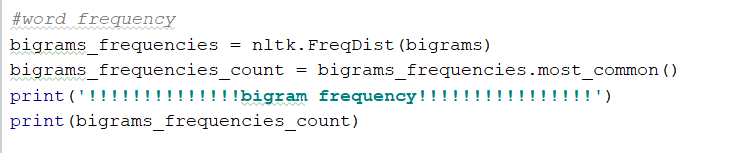


**Output**



**Word frequency of the bigrams**

**Code**

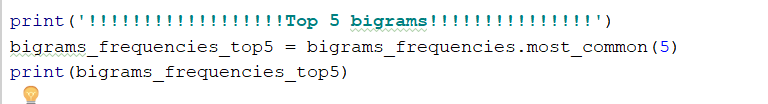


**Output**

https://github.com/sirisha1206/Python/raw/master/Lab/Lab2/Documentation/3q_o2.PNG

**Top 5 frequent bigrams**

**Code**

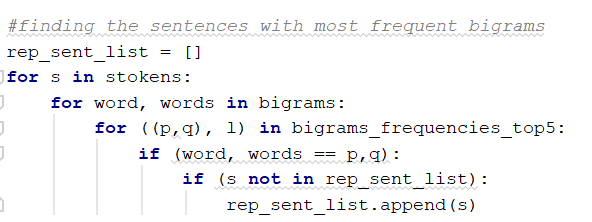


**Output**

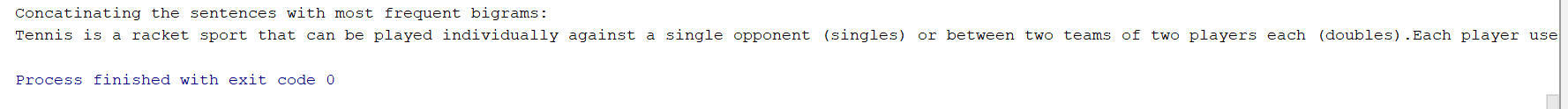
https://github.com/sirisha1206/Python/raw/master/Lab/Lab2/Documentation/3q_o3.PNG

**Sentences with top 5 bigrams**

**Code**

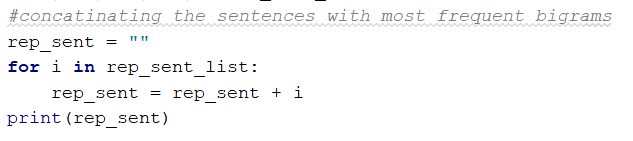


**Output**

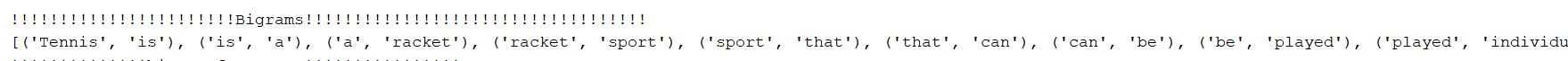


**Concatenation of sentences with most frequent bigrams**

**Code**



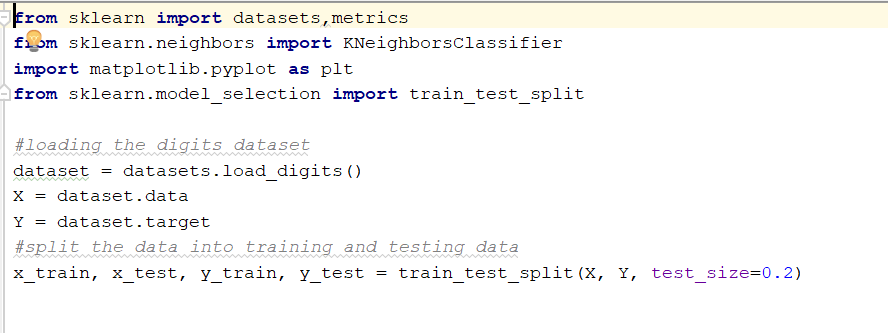
**Output**



**Task 4**

K Nearest Neighbor Algorithm

For KNN algorithm we have used digits dataset from sklearn library.we have split the data into training and testing data.



Steps to get the accuracy of KNN with different k values:

1.Select the k range from 1 to 60

2.Create an object for knn classifier with varying number of neighbors.

3.Train the model with the training data

4.Predict the value

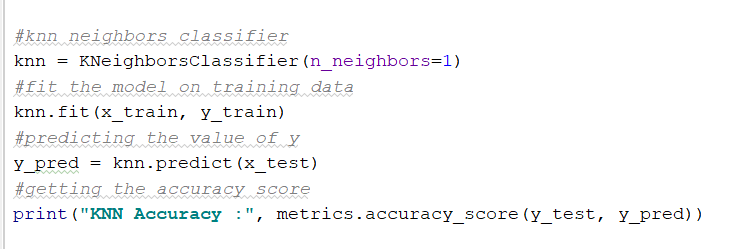
5.Get the accuracy

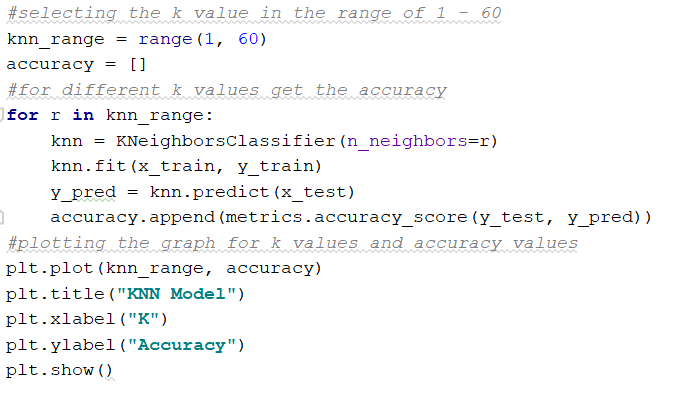
6.Plot the graph

**Report**

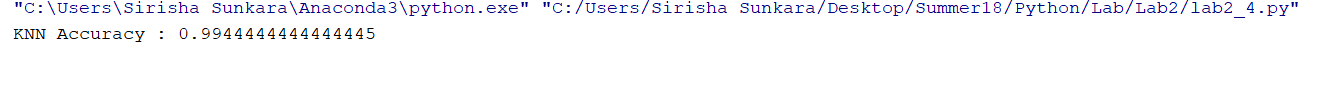
As the value increases the accuracy score is decreasing which means they are inversely proportional.When the k value is small,it has high variance and low bias.When the k value increases it has low variance and high bias with smooth boundaries.

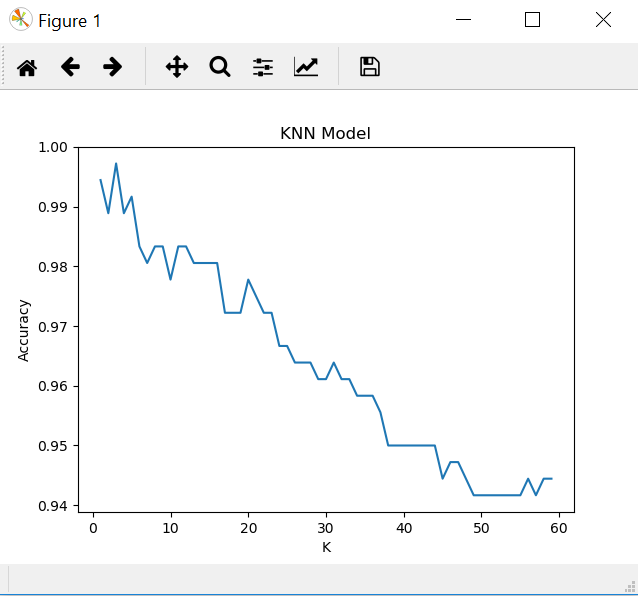
**Code**





**Output**





**Conclusion**

All the given tasks have been implemented

[Source Code](https://github.com/sirisha1206/Python/tree/master/Lab/Lab2/Source): https://github.com/sirisha1206/Python/tree/master/Lab/Lab2/Source

[Video Link](https://youtu.be/Ps31tqm2XJM): https://youtu.be/Ps31tqm2XJM

**References:**

<https://www.stackoverflow.com>

[https://www.geeksforgeeks.com](https://www.geeksforgeeks.com/)