Kevin Thomas

CSE 6361-001

1001544593

PROJECT -2 REPORT

Structure of the Code

Packages imported are,

Os ,Counter, time ,nltk, Porter Stemmer, word_tokenize

1) First download the data set from the given URL.
2) I am creating a dictionary to store the value of the dataset,
3) Then I split the data into train and test datasets (i.e) the first 500 will be in the train and the remaining 500 will be on the test data.
4) With the given data we have to clean the data by removing unnecessary words by using various methods like set(which gives the unique words in the list) and other functions
5) Then we find the probability of each class and a word in each classes.
6) With the given probabilities we implement the Naive Bayes classifier and find the accuracy which is the avg of all the classes.

Result:

From the given data the accuracy and the execution time can be got.

Accuracy: 76.89537907581516 or 76.89%

Excecution time: 273.9339377252836 seconds

To run the code , it can be ran on any python IDE.