**Name:** Akshya Srinivasa Raghavan

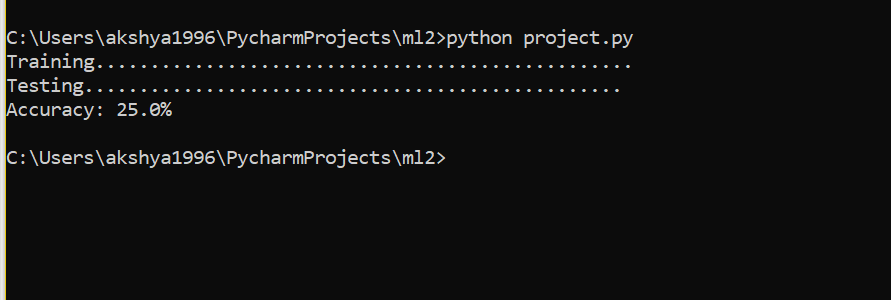
**Id:** 1001547268

**Project:** Bayesian learning for classifying netnews text articles

**Programming Language used:** Python

**Compilation Steps:** python project.py

**Output:**



**Code Explanation:**

* Half of the data is train and the other half is test. Since we have 1000 documents, I have set the number of train data to be 500 and the number of test data to be 500
* Data\_folder is set to the folder that contains the data
* First, we list the folders in this data folder, and we traverse each folder with this and each file within each sub folder
* We open each file and replace the symbols with a space
* We then split the data into words using space as a delimiter and store them in a list
* We find the frequency of each word in this list and store them in a dictionary
* This is done for the first 500 data
* Now we do the same thing, but list the file folders in the reverse alphabetical order to consider the last 500 data for testing
* We have the list of words in the test data
* We then find the probability of each word in the list for test data with that of the final dictionary from train data and then the classes is found
* We then print the accuracy

The accuracy according is **25.0%**