CS 6375

ASSIGNMENT \_\_\_\_\_4\_\_\_\_\_\_\_

Names of students in your group:

Pranathi Peri(pxp162530)

Nithin Vasireddy(nxv160230)

Number of free late days used: \_\_\_\_\_\_0\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

**Read Me File**

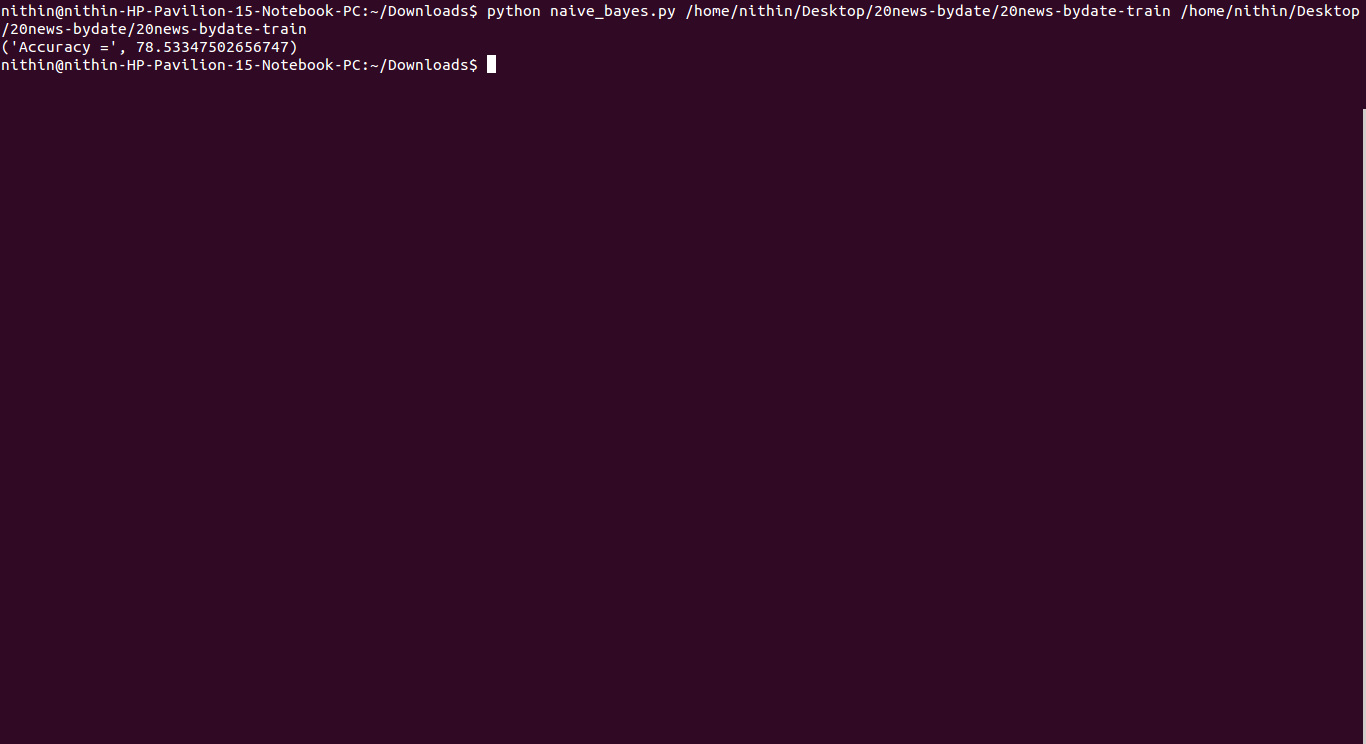
**Language used:** Python

**Packages Used:** Collections, OS, Sys, Math, re

**Steps to run the Code:**

1)Run the file naive\_bayes.py which is the file where the learner is implemented. As arguments to this file give the location of training and testing root folder.

**Screenshots:**



We have implemented the multinomial Naïve Bayes Classifier.

Assumptions and Analysis:

1. We have excluded the lines from the start of file upto lines from the starting with “Lines:XXX”
2. We have included the stop words in our model.
3. From the data set we have tokenized the words by having a regular expression which handles special characters.
4. The algorithm uses Laplace smoothing.
5. We have done the calculations in log-scale to avoid underflow.

Results:

1)When we consider the five classes sci.electronics, sci.med, sci.space, soc.religion.christian and talk.politics.guns. The accuracy was 91.67

2)When we consider the five classes alt.atheism, comp.graphics, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, comp.sys.mac.hardware.The accuracy was 78.53