**ASSIGNMENT 2**

**MACHINE LEARNING**

**SAKET DINGLIWAL 2015CS10254**

**PART 1**

**NAIVE BAYES**

**Question 1**

Naive Bayes was implemented without using any cleaning of the text data ( simple string.split() ) was used to tokenize the data.

The training accuracy obtained was - 68.844

The test accuracy obtained was - 37.388

**Question 2**

Test set accuracy for random selection of class, with each class being equally likely was obtained as - 11.188

The accuracy for choosing the most frequent class was obtained to be – 20.088

The test set accuracy for the naive bayes was 37.388

which is almost double than that by majority selection and triple than that by random selection.Although this is not much ahead of these simple baseline, but good enough for the simplistic naive bayes model.Since there are 8 classes, the random one was expected to be close to 1/8. Since there are uneven number of documents with each classes, the majority is around 20%

**Question 3**

4726 0 4 34 9 33 0 216

2028 2 4 56 13 44 0 155

1958 1 28 140 24 120 0 270

1674 2 4 241 62 218 1 433

656 0 0 55 91 436 0 1069

616 0 2 33 34 469 2 1694

435 0 0 9 16 210 0 1674

975 0 2 8 13 210 1 3790