Assignment 1-B Naive Bayes Algorithm

A)

___We have implemented using the algorithm with the major usage of Bayes theorem,
Bayes theorem states that :
: P(class|data) = (P(data|class) * P(class)) / P(data)
Where here we know the data(beforehand) and we have to calculate the probability of it being belonging to a particular class.
In our example, we have 2 classes only.

1) Spam
2) Not Spam.

So in the formula stated above, effectively we have to calculate the RHS for our answer, and we notice that after opening the right hand side of the formula, thata denominator is common for all terms and we only need to calculate our answer based on the numerator.

In our dataset we have implemented the Naive Bayes algorithm as follows.

1) We calculate the count of text occurrence with value either yi =0/1, let this be **[DENOM]**
2) And then we calculate the count of occurrence of the words in the text with yi =0/1 and count of words **[NUM]**
3) Finally, our answer becomes maximum of all the words taken into consideration
   I.e maximum of **[NUM]/[DENOM]** for all the words in the given training set.
4) And we finally associate the given text with spam/not spam with the output of step 3.

B)

Each fold we divide our dataset into 1:6 ratio and then we take this 1 ratio as our test data set and rest of the dataset to train our dataset.
The accuracy of our model over each fold is shown below.
The overall accuracy of our model is also shown below:

```
Fold # 1
The accuracy is  : 79.8 + 0.0
Fold # 2
The accuracy is  : 79.1 + 0.6999999999999957
Fold # 3
The accuracy is  : 79.56666666666666 + 0.8730533902472503
Fold # 4
The accuracy is  : 78.57499999999999 + 1.876665926583633
Fold # 5
The accuracy is  : 78.39999999999999 + 1.7146428199482258
Fold # 6
The accuracy is  : 78.74999999999999 + 1.750000000000001
Fold # 7
The accuracy is  : 78.39999999999999 + 1.833030277982337
Overall accuracy is  : 78.39999999999999 + 1.833030277982337
```

C)

Disadvantages of Naive Classifiers:

a) Correlated features might affect our model.

b) In text classification, it doesn’t take into account the sentiment of the text, and just works on the basis of probability(just mathematical) and is hugely dependent on the training dataset we provide.

c) The algorithm might face the data which have zero frequency, and this can be handled by using laplace smoothing/correction.