SPAM CLASSIFIER:

The steps involved are as follows to implement the spam classifier by Naïve Bayes Algorithm:

1. The training data-set was obtained from the internet(*emails.csv*)
2. Selected all the valid words (*features*)from these emails and created a dictionary
3. Converted all the training mails into vectors (which denotes the number of occurrence of words from our dictionary)
4. Calculated the Prior
   Probability that an email is spam in given training dataset
   p_hat= number of emails that are spam/ total number of mails

   Probability that an email is non spam in given training dataset
   =(1-p_hat)
5. Calculated the likelihood for all the words in the dictionary
   Likelihood (word i/spam) =number of occurrences of word in the training set when mail is spam/total number of word occurrences in spam mails
   Likelihood (word i/non spam) =number of occurrences of word in the training set when mail is not spam/total number of word occurrences in non-spam mails
6. Used Naïve Bayes function which takes an email as input and returns 1 (SPAM) or 0 (NON-SPAM) based on the Posterior Calculation
   Posterior (spam/word)=p_hat * /\(likelihood for all words/spam)
   Posterior (non-spam/word)=(1-p_hat) * /\(likelihood for all words/non-spam)
7. Based on the comparison of values if Posterior (spam/word)> Posterior (non spam/word) Then the email is classified into spam category otherwise non-spam

   Testing
8. from the test folder we are reading the txt files (files ending with .txt) and storing them into csv file (test.csv)
9. for each of the mails reading from the test.csv file we trigger the naïve bayes function and store the result into a label
10. print the label