Assignment 3: Naïve Bayes algorithm for text classification.

First part:

In this assignment, we will redo the task of classifying documents (assignment 2) using the same Reuter dataset. But this time, you should implement the multinomial naive Bayes algorithm instead of KNN. Naive Bayes used to be the de facto method for text classification. Try various smoothing parameters for the Naive Bayes learner. What's the accuracy of your learner? Which parameters work best?

Second Part:

In this part, you will compare between the performance of k-NN classifier and Naïve Bayes classifier for text classification. Follow the steps below:

1. Take the best classifier from your second assignment (k-NN). Chose the best value of k and best measure of distance/similarity that gave the best performance.
2. Compare the best k-NN with Bayesian classifier. Run 50 times both the k-NN and Bayesian learner. Compute mean and standard deviation of the results. Then, compute t-statistic and at significance levels of 0.005, 0.01, and 0.05 compare which algorithm (k-NN or Bayesian) is better. Report the results in a paper and submit it.