ASSIGNMENT 1: NAIVE BAYES CLASSIFIER

A) Accuracy for local implementation (in percentage) on variety of datasets:

Accuracies on different dataset (For full data):
Training data: 16.7543
Development: 22.1023
Test data: 13.9843

Accuracies on different dataset (For small data):
Training data: 14.7543
Development: 12.1023
Test data: 10.9843

B) Calculation of total nos. of parameters:

Every word has its own Hash map in the local implementation. Hence total number of parameters used to implement naive Bayes are :
We have 50 labels and for each label, we have two parameters as Prob. (Data X = x| Label Y = y) and Prob. (Y= y) where X represents the data and Y represents the labels.
Hence total no. of parameters = Total nos. of words * (Total no. of labels) + (Total labels – 1).

C) Running time analysis

Total running time (for training data of full size) = 83.54 milliseconds
Total running time (for test data of full size) = 20.47 milliseconds.
Total running time (for training data of small size) = 15.78 milliseconds
Total running time (for test data of small size) = 3.64 milliseconds.
Total running time (for training data of full size) = 83.5 milliseconds
Total running time (for test data of full size) = 20.47 milliseconds.