ML A2 (Naive Bayes Classifier) Report

Part c

The features tried to improve accuracy are as follows:-

- Stopword removal :- Stopwords were removed using nltk library. This showed a minor overall decrease in accuracy.
- Word Stemming :- Words were converted to their stemword before processing. This also resulted in slight decrease in accuracy over public test set.
- Bigrams:- I tried to include bigrams in the bag of words for every review. This did not show any improvement in classification.
- Term overall count:- I maintain a count of the number of times a word is seen in all the reviews given to us. Using this i divide all the log scores of all the words by log of this count. This gives a slight improvement in overall accuracy of the Naive Bayes model.