2)Naïve Bayes :

Approach:

* The preprocessing includes cleaning of data by removing punctuations
* We used k-fold cross validation to train and test
* Classify the points in test set based on word frequencies in train set
* Finally compute the mean and standard deviation

Smoothing:

* Laplace smoothing was performed in order to avoid 0 probabilities by adding a smoothing parameter: 1/N .where N is the number of words in the vocabulary which is 3543.

Dataset:

* The dataset used is a collection of user reviews in string format followed by their classification as good or bad(0 or 1)
* <a1_d3.txt>

Dependencies Used:

* Numpy for working with array operations efficiently
* Math library as it contains log function
* Random library for shuffling the data

Results:

* f1\_score: 0.7934845775757348 +/- 0.013018017529589032
* accuracy: 0.794 +/- 0.0058309518948453055
* [Finished in 3.9s]
* (for 5-split cross validation)