Reply to Questions in Q3a:

1. We split the train set into two parts which are sub-train and sub-test sets, respectively. The split ratio of the sub-train set to the sub-test set is 8:2. The mean accuracy of the sub-train set is nearly 1 and the mean accuracy of sub-test set attains 0.99.

2. There is a parameter which is called alpha in the Naive Bayes classifier for multinomial model. This is a smoothing parameter. If it is zero, then there is no smoothing. We adopt the GridSearchCV to automate the search of this parameter, alpha. In our case, the optimal parameter, alpha, is found as 3.8776.

3. The model is based on the Nave Bayes algorithm. The underlying principle is based on Bayes' rule. Here Nave Bayes classifiers assume that the features are independent of each other with the given the class label.

   Indeed, the probability of the tags given the text is related to the probability of the tag in the data set and the probability of features conditional on the tag. For example, the tag "football" can be predicted by the given text. If there are some wordings accompanying the same tag label in the train set, then the text with those wordings is more likely to relate to that tag label. To be more specific, consider a text with the words "footaball", "worldcup" and the name of a football star, it is more likely to be predicted as "football" tag.