**Question1**

the only problem I had countered in data that it was full of duplicated records (more than half of records).

I used a simple method to just drop duplicated with keeping the last values in order not to lose a lot amount of data which is not good for the model.

**Question2**

The reasons I choose Multinomial Naive Bayes are as follows:

* It doesn’t require as much training data (our data has reduced to about the half because of removal of duplicates)
* It is fast and can be used to make real time predictions (which in our case is suitable to implement on a web)
* It is easy to interpret with text data (it is essential for our case as all our data is considered text)
* It is easy and simple to implement

**Question3**

I used stratify parameter while splitting data in order to make sure that all labels are represented in both training and testing sets.

**Question4**

I think if we increased number of data points we will get a better model with much better accuracy

**Question5**

I used accuracy, F1, recall and precision scores to evaluate the model and give me a good intuition on how it works.

**Question6**

* The model accuracy is not much good due to low amount of data.
* The mistyping of user into web can lead to bad results so preprocessing is required to the raw data entered by user before prediction.