**Part D:**

If we are \*\*not removing\*\* the stopwords from the text, the accuracy and F1 score are better compared to the text where the stopwords are \*\*removed\*\*. Why does this happen? When the stopwords have been removed the likelihood of the other words in the training data remains the same because the stopwords are independent of the other words in the training data. Now, during the testing, since there are no stopwords in the vocabulary the logarithmic addition will result in less posterior value compared to the posterior value when stopwords are present. This is happening because of stopwords appearing in the respective classes are removed.

**Part E:**

The fundamental drawback of the naive Bayes classification algorithm is that negation phrases like "doesn't" and "haven't" are ignored by the algorithm. To make our algorithm understand the difference between does and doesn't we can concatenate not with the doesn't (e.g.: don’t will be not\_don’t), then our algorithm will take account of these words when we are calculating the likelihood, which will result in improving recall and precision.

**Bonus:**