TASKS

Task 1- Done
Task 2- The accuracy is not impacted that much. Actually, in this case, the removal of stop words actually makes the accuracy of this model worse.
Task 3- Done
Task 4- Features that have to do with the dependencies between words could help with text classification. For instance, the ambiguities of words does not get ruled out when you use naive Bayes. If you have a word such as lead, this has both a use as a noun and as a verb. Thus, this confusion could lead to a incorrect classification of text. Additionally, you could take into account part of speech as a feature of text classification. For instance, some authors may be more inclined to use some parts of speech more than other parts of speech. Thus, it would be easier to classify the difference between two sets of text that are written by two different authors. These features are not completely independent of the Bayes 'bag-of-words' model. For the dependencies on words, the dependencies are still driven by the actual distribution of words in the sentences and for the parts of speech feature affecting classification, that classification process is also driven by the actual distribution of words within the text. This is because there are certain words that only have one part of speech that can be assigned to them.

RESULTS

MULTINOMIAL BAYES CLASSIFICATION

[INFO] Fold 0 Accuracy: 0.765000
[INFO] Fold 1 Accuracy: 0.850000
[INFO] Fold 2 Accuracy: 0.835000
[INFO] Fold 3 Accuracy: 0.825000
[INFO] Fold 4 Accuracy: 0.815000
[INFO] Fold 5 Accuracy: 0.820000
[INFO] Fold 6 Accuracy: 0.835000
[INFO] Fold 7 Accuracy: 0.825000
[INFO] Fold 8 Accuracy: 0.755000
[INFO] Fold 9 Accuracy: 0.840000
[INFO] Accuracy: 0.816500

MULTINOMIAL BAYES CLASSIFICATION W/O STOP WORDS

[INFO] Fold 0 Accuracy: 0.765000
[INFO] Fold 1 Accuracy: 0.825000
[INFO] Fold 2 Accuracy: 0.815000
[INFO] Fold 3 Accuracy: 0.830000
[INFO] Fold 4 Accuracy: 0.795000
[INFO] Fold 5 Accuracy: 0.830000
[INFO] Fold 6 Accuracy: 0.835000
[INFO] Fold 7 Accuracy: 0.835000
[INFO] Fold 8 Accuracy: 0.760000
[INFO] Fold 9 Accuracy: 0.820000
[INFO] Accuracy: 0.811000

BINOMIAL BAYES CLASSIFICATION

[INFO] Fold 0 Accuracy: 0.805000
[INFO] Fold 1 Accuracy: 0.840000

[INFO] Fold 2 Accuracy: 0.835000
[INFO] Fold 3 Accuracy: 0.825000
[INFO] Fold 4 Accuracy: 0.835000
[INFO] Fold 5 Accuracy: 0.825000
[INFO] Fold 6 Accuracy: 0.845000
[INFO] Fold 7 Accuracy: 0.835000
[INFO] Fold 8 Accuracy: 0.790000
[INFO] Fold 9 Accuracy: 0.855000
[INFO] Accuracy: 0.829000