IST 707 Applied Machine Learning
HW8: Lie Detection and Sentiment Classification with Text Mining
Name: Jainish Savaliya

Model	Parameter Settings	Accuracy Lie Detection	Precision Lie Detection	Recall Lie Detection	Accuracy Sentiment	Precision Sentiment	Recall Sentiment
SVM	degree = c(1, 2, 3), scale = c(0.001, 0.01, 0.1, 1.0), C = seq(0.1, 2, length = 10)	Train:82.61 Test:64.17	0.5833	0.70	Train:95.65 Test:97.78	0.962	0.956
Naive Bayes	use kernel = c(TRUE, FALSE), laplace = c(0, 1), adjust = c(0,1,2)	Train:55.56 Test:75	0.625	0.98	Train:60.87 Test:51.11	0.511	0.98

At first I read the txt file into csv format the with separator being the tab. And had a look at the glimpse of the dataset and it had three variables lie, sentiment and review. So, In the data cleaning process I made sure that the variables were factorized. Fake or true for lie and positive or negative for the sentiment.

Then I tokenized the reviews to make the tokens out of the sentences and removed all the stop words and dropped all the spaces out of the dataset using filter and antijoin. Finally, I lemmatized all the tokens that I created,
Relocated index variable and created the data partitioning and removed the index before training the model.

For the SVM training model for both lie and sentiment detection I used the tuning parameters given in the table above. the reason to choose those parameters in addition of using 5-fold cross validation were to optimize the models complexity and trade off between accuracy and generalization and I used polynomial kernel for that.

For the Naïve Byes training model for both lie and sentiment detection I used the tuning parameters given in the table above and the reason to choose those parameter in addition of using 5 fold cross validation was to smoothen the probabilities through kernel, Laplace and adjust parameter. Finally I chose the parameters give in the table to maximize the performance.

If we Compare the Performance of the SVM and Naïve Bayes for Sentiment and Lie detection sentiment prediction is easier task compared to Lie detection. As we can see for both the models nearly sentiment task performed better than lie. As per my understanding overall both the sentiment classification and lie detection are challenging tasks, having said that lie detection is more difficult due to the complexity of analyzing language, context and human behavior to identify deception.