David Saffo
Comp 379
11/12/2016

Pre-Processing:

The data came in two text folders label pos and neg. To start I read each of these files into their own lists then merged them together with pos reviews on top and negative reviews on the bottom. I then added a second dimension to the list by adding pos or neg labels for each review. Once this was done I shuffled the indexes around and split the data 70, 15, 15. Looking at the data I wanted to read in each review as a new line and break them down by words later.

Training/Development:

The model was trained reading reviews in one at a time taking their labels and looking at the words and word frequency. When testing the model with the development set I chose not to remove words from a stop list as it seemed to lower performance and instead found that removing words that were less than 2 characters long provided the best results.

Results and Metrics:

Predictions were made by getting a probability score of the review being pos vs neg and the bigger of the two being the prediction. After tuning with the dev set the test set scores about 76-77% accuracy and achieves similar f1score. Since the accuracy score and f1 score are so close I would say it does not really matter which one you choose for this instance. However, in the general case f1 score gives a bigger picture of the results accuracy.