Progress Report

Sunday, November 29, 2020   10:45 AM

As part of the final project, I have opted for the classification problem. Below is the status for my progress:

1) Progress made thus far:
    a. I have done the preprocessing of the data like, removing stopwords, any character which is not alphanumeric, remove punctation.
    b. I have tried various model, Navies Bayes, Logistic Regression, SVM, however with all of these I was not able to beat the baseline. I am still revolving around .66 -.70
    c. I have tried Neural networks as well but still not able to beat the baseline. I am still revolving around .66 -.70
    d. I have tried CNN with various filters and kernel but still not able to beat the baseline. I am still revolving around .66 -.70
    e. I have tried LSTM with various combinations like, Bidirectional, different units (64,128,256) and also regularization but still it revolve around .68 -.71

2) Remaining tasks:
    a. I am still working on improving my vocabulary as well as will try BERT to see if I can beat the baseline.

3) Any challenges/issues being faced:
    a. I am still trying to find on how to improve the model F1 score to beat the baseline. I think improving the vocabulary will be helpful but not sure what other things should I try.