

CS 410 Progress Report

1. Which tasks have been completed? I have completed the task of preprocessing the data. I have also completed the task of implementing different models for the Text Classification Competition. I have tried using SVMCNN, CNN, and LSTM. I have also experimented a bit with logistic regression, RBF kernel SVM, and different combinations of those models. The baseline score has not been beat yet.
2. Which tasks are pending? Currently I am trying to fine-tune a BERT model on my preprocessed data. We will see if this beats the baseline or not.
3. Are you facing any challenges? Only with beating the baseline. The best F1 score achieved so far came from feature selection from a CNN trained on context data and TFIDF vectorization of the response data with LASSO to determine useful words. These features were combined and fed into a linear SVM.