CS 410 Progress Report

1. Progress made thus far

We have already implemented the BERT-based pre-trained model using python and then using linear transformation with train and validation set to fine-tuned the model for text classification. We've already finished the majority of our code but got F1 score of 70.14, which is below the baseline.

2. Remaining tasks

We still need to make improvements to our code: adding CNN or doing data cleaning for our training and test set. In addition, we still need to make full documentation for our code and optimize our code if possible after beating the baseline. Also, we need to make our own tutorial video at the end to show our functionality and code.

3. Any challenges/issues being faced.

Currently, we can not simply improve our F1 score by adding epoch or simply changing hyperparameter like learning rate, l2 regulation rate, or batch size to improve our performance. As a result, we need more efficient and effective method to increase our model's precision and recall.