Project progress report

1) Which tasks have been completed?

Firstly, after reading through the paper Learning Word Vectors for Sentiment Analysis, I found the database used by the paper (https://ai.stanford.edu/~amaas/data/sentiment/). I downloaded the database and separate the data into testing data and training data.

Then I start to construct the skeleton of my coding part. I create three python files. One is classifier.py which will mainly be in charge of n-gram and other classification methods implemented. One data_loader.py file used to read data to my classifier. The last one is the main.py file which will parse the argument and run the program.

I also finished implementing a unigram and bigram classifier for the project. I also add Laplace smoothing method to the model to avoid overfitting and overflow.

2) Which tasks are pending?

The classifier still needs improvement to more gram and smoothing.
I will need to work on data_loader in the future to ensure the data passed into the classifier is valid.
The whole implementation of the web application is still pending. I will start working on it after finish implementing the classifier.

3) Are you facing any challenges?

It seems like the data I got is bigger than I expected. Maybe I will need to modify the size of it if the size is affecting the performance.