Write up for assignment 3 (NAME: Shekhar Kumar Yadav, ID: 19310032)

- I have first created a file containing all the tweets (train and test) from the train.txt and test.txt files using the attached python codes of 'train_clean.py' and 'test_clean.py'. The generated file is saved as 'train_test_cleaned.csv'. The first 15130 lines are train tweets and the rest 1868 tweets are test tweets.
- I have also extracted the labels of all the corresponding tweets and (train and test) and the generated file is saved as 'sentiment_cleaned.csv'. 'Positive'→0, 'Neutral'→1 and 'Negative' →2. The first 15130 sentiments are for training and the rest 1868 sentiments correspond to the test tweets.
- I have used Doc2vec to assign every tweet a vector and have used that vector along with known labels to train a neural network
- I have used a keras 'Sequential' model with a flatten layer and two Dense layers of sizes 128. The final dense layer is of size 3 for classification. The optimiser is 'Adam' and the loss function is 'sparse_categorical_crossentropy'. The epoch set at 100.
- Once the network is trained, the test cases are assigned a label and accuracy is calculated using the known label.

The main file in sentiment_analyzer.py

The accuracy of the model is 42%