What the .zip file contains:

1. This readme.docx
2. 3 Jupyer notebooks attached.
   * Code for Submission.ipynb. This is the main notebook where you should start with.
   * GridSearch.ipynb. This notebook contains the codes to run GridSearch to tune the parameters for better accuracies.
   * Baseline Accuracy and NN-Original.ipynb. This notebook calculates the baseline accuracies for non-processed and processed corpus. It also contains the codes for the neural network models tested on our dataset.
3. 2 .csv files and 1 excel file are included for all the data:
   * ted\_main.csv (this is directly from Kaggle)
   * transcripts.csv (this is directly from Kaggle)
   * processedfeaturesDF.xlsx (this is generated so that pre-processing need not be done every time the code reruns)
4. Lastly, there would be a .py file attached, textpreprocess.py which contains the functions to normalize, lemmatize, stem etc. words in the corpus.