This repository contains various files and folders but for viewers the important folders are the four folders namely .py files, Text Files, DataFrame and Output.

Information regarding each folder is given below, please go through this.

**.py files**

* This folder contains .py files used to extract the text from the article and also files used to perform text analysis on the text files.
* The .py file (scrape.py) is used to extract the title and text from the URL provided in Input.xlsx in assignment. While doing so the title and text are saved in the .txt files with URL\_ID as the name of .txt file in folder Text Files.
* DataFrame.py file is used to create Article.csv file in which text from single article is stored in single row.
* NLP Model.py is used to create tokens from and clean text stored in DataFrame.py
* Text Analysis.py is used to perform sentimental analysis over the DataFrame.csv and also compute various variables mentioned in Objective word file.

**Text Files**

* In this folder all the text files are saved which contains title and text scraped from the article related to given URL with URL\_ID as name of text file.

**DataFrame**

* This folder has .csv files which are used as raw files over which several operations are performed. Not necessary to be viewed but are important in building NLP model.
* This folder has the .csv files which are used to save the content of each text file from folder Text Files as one row, so that we can perform sentiment analysis, tokenization and cleaning on each row.
* This folder also contains .csv files in which clean words, positive words, negative words and total words are saved in order to perform text analysis on each row.

**Output**

* This folder contains Output Solution.csv file which can be considered as output or result of the assignment. This file has all the variables computed and stored in respective feature.