**README FILE**

**FILES INCLUDED IN THE FOLDER:**

* Tweets.csv – Dataset
* ClusTop.py – Source code file
* run – executable file

**STEPS TO EXECUTE:**

1. Download all the files from GitHub repository.
2. Install python 3.7 version
3. Open command prompt and go to the path of location of downloaded files.
4. Install all the required packages

Commands used to install packages:

* pip install pandas
* pip install matplotlib
* pip install numpy
* pip install sklearn
* pip install nltk – you need to download few more packages from nltk. So, follow below commands to

>>python

>>import nltk

>>nltk.download()

1. Now execute the “run” batch file
2. Once, you have executed the “run” file then you will be getting the user interface of “ClusTop”.
3. First click on “Upload Tweet Dataset” and select “Tweets” csv file from the “Dataset” folder and then click on open.
4. For preprocessing the dataset, click on “Preprocess Dataset”
5. Once preprocessing is done, Run “ClusTop-Word-NA Algorithm”. You will get the tweets and detected topics by using this algorithm (CoWord).
6. If you want to check for Bigrams, click on “Run ClusTop-BIG\_NA Algorithm”. You will get the tweets and detected topics by using Bigram Algorithm.
7. To check the LDA Algorithm output, click on “Run LDA Algorithm”. (Wait for one-two minutes to get the result)
8. To comparison the results of both ClusTop and LDA, you can click on “Comparison Graph” and “Comparison Table”.

**NOTE:** We need to run the algorithms one by one

1. After preprocessing, click on “ Run ClusTop-Word-NA Algorithm”
2. Then click on “Run ClusTop-BIG\_NA Algorithm”
3. Then click on “Run LDA Algorithm”
4. At last, click on “Comparison Graph” and “Comparison Table”.

As all the operations are interlinked in the designed UI, We must follow the above order.