**Make sure before running the Tweet\_Cluster\_K\_Means.py**

1. To download/create InitialSeeds.txt file which contains k initial selected cluster centroids.

http://www.utdallas.edu/~axn112530/cs6375/unsupervised/InitialSeeds.txt

1. A real-world dataset sampled from Twitter during the Boston Marathon Bombing event in

April 2013 that contains 251 tweets. The tweet dataset is in JSON format and can be downloaded from

http://www.utdallas.edu/~axn112530/cs6375/unsupervised/Tweets.json

**Compile and Run Instructions**:

**Commands to run**:

--> Tweet\_Cluster\_K\_Means.py <numberOfClusters> <initialSeedsFile> <TweetsDataFile> <outputFile>

--> numberOfClusters is optional, if not given then default value 25 will be taken

Tweet\_Cluster\_K\_Means.py <initialSeedsFile> <TweetsDataFile> <outputFile>

**Example commands**:

1. python Tweet\_Cluster\_K\_Means.py 25 InitialSeeds.txt Tweets.json tweets-k-means-output.txt

2. python Tweet\_Cluster\_K\_Means.py InitialSeeds.txt Tweets.json tweets-k-means-output.txt

**Results:**

After running Tweet\_Cluster\_K\_Means.py a file tweets-k-means-output.txt is created in your current directory which contains the results after clustering the tweets based on the initialSeeds.txt containing k initial selected cluster centroids.