Report

This code implements the k-means classification methods to classify tweeter data into 25 categories based on Jaccard Distance. The code is running under Jupyter Notebook. The code should be in the same folder with InitialSeeds.txt and Tweets.json files. Call the tweets\_k\_means() function with default parameters to export results. The default parameters are set as below:

tweets\_k\_means(K=25, initials="InitialSeeds.txt", tweetsFile="Tweets.json", output="tweets-k-means-output.txt")

User could customize the “initials” and “tweetsFile” for classification purpose.

The results are saved in “tweets-k-means-output.txt” file by default.

The final SSE value is listed at the last line of the export txt file.