### INTRODUCTION:

In this assignment, K-Means is implemented which is an example of Unsupervised learning. The tweets are classified into clusters based on the Jaccard distance of the tweet to every cluster center. The Jaccard distance is calculated as the difference of the sizes of the union and intersection of the two tweets divided by the size of union of the tweets. We assign the tweets to the nearest clusters (distance) and once all the points are assigned to the respective clusters, the cluster centers are recomputed again and the tweets are assigned to the nearest cluster. This process goes on until the centroids converge. The sum of the squared error is calculated as the sum of the square of the distance of each tweet from its cluster center.

### RESULT:

The K-Means algorithm is successfully run and the corresponding distribution of tweets to respective clusters and the sum of squared error are calculated. The clustered tweets and SSE value are written to an output file. The following images show a snapshot of the output file:





### CONCLUSION:

Thus, K-Means clustering algorithm is implemented for classifying various tweets and SSE is calculated.