clustering.py: This file is used to generate clusters using k-means clustering algorithm

Libraries and classes used for clustering:

from sklearn.cluster import KMeans

please simply run the cluserting.py file, it will 3 clusters in the same folder.

logisticregression.py: This file is used for apllying logistic regression approach.

Libraries and classes used for Logistic Regression:

from sklearn.linear_model import LogisticRegression

please simply run the logisticregression.py file, it will generate predicted_tweetslr.txt file in the same folder.

svm.py: This file is used for applying support vector machine approach

please simply run the svm.py file, it will generate predicted_tweetssvm.txt file in the same fold

Libraries and classes used for svm:

from sklearn import svm

from sklearn import cross_validation

from sklearn.grid_search import GridSearchCV

from sklearn.model_selection import GridSearchCV

In this svm 10 and 20 folder cross validation is formed

tweets.txt: This file contains labelled tweets for training purpose.

unlabelled_tweets.txt= This file contains testing tweets for testing purpose. Tweet in this file is in the format (class_label, tweet_text), but for the testing I have generated X of tweet_text only.

predicted_tweetslr.txt= This file contains predicted tweets using LR. This file contains line in the format (predicted_label, original_label, tweet_text)

predicted_tweetsvm.txt= his file contains predicted tweets using SVM. This file contains line in the format (predicted_label, original_label, tweet_text)