

## Read Me

The folder contains the following files :

1. sourceCode.py – source code in Python programming language.
2. Input Files:
  1. rawTweets.txt – Set of Raw tweets extracted from json-file.
  2. bow\_class{A|B|C|D|E|F}.txt – Initial set of Bag of words, selected manually. Characters represents class, explained below.
  3. acronym.txt – Contains the list of Acronyms used.
  4. stopWordddList.txt – Contains the set of stop words.
  5. trainingDataSet.txt – 20% labelled dataSet used for training SVM.
  6. manualClassifiedFromFIRE.txt – Text file containing all the labeled data from FIRE.
3. Output Files:
  1. Plots:
    1. System\_Accuracy.png
    2. Class\_Resources\_Available.png
    3. Class\_Resources\_Required.png
    4. Class\_Medical\_Rsc\_Available.png
    5. Class\_Rsc\_at\_Locations.png
    6. Class\_Medical\_Rsc\_Required.png
    7. Class\_NGO\_Activites.png
    8. Class\_Infr\_Damage\_Restoration.png
  2. System Classified Tweets.
    1. Classified\_{A|B|C|D|E|F}.txt – List of TweetIDs classified into their respective classes by the system. Character meaning explained below.
    2. Classified\_Neutral.txt – List of TweetIDs not classified into any of the above classes.

Representation of characters used above:

Each character describe a particular class, like

A – *What Resources were Available*

B – *What Resources were Required*

C - *What Medical Resources were Available*

D - *What Medical Resources were Required*

E - *What were the requirements/availability of resources at specific locations*

F - *What were the activites of Various NGO's/Government Organizations*

G – *What Infrastructure damage and restoration were being reported*

Note: All input/output files should be kept / will be generated in the same working directory.