

I stand with you: using Emojis to study Solidarity in Crisis Events

By

Priyam Mukherjee (GCECTB-R16-2019)

Shawan Basu (GCECTB-R16-2026)

Suvam Das (GCECTB-R16-2034)

Defining Solidarity

We start by defining what we mean by solidarity. The concept of solidarity has been defined by scholars in relation to complementary terms such as “community spirit or mutual attachment, social cooperation or charity”. In our work, we use the definition of expressional solidarity, characterized as individuals expressing empathy and support for a group they are not directly involved in (for example, expressing solidarity for victims of natural disasters or terrorist attacks).

Need for Solidarity Analysis

The collective enactment of online behaviors, including prosocial behaviors such as solidarity, has been known to directly affect natural disasters and social movements. Social media, due to its increasingly pervasive nature, permits a sense of immediacy **[a notion that produces high degree of identification among politicized citizens of the web, especially in response to crisis events]**

- We study how emojis are used to express solidarity in social media in the context of two major crisis events - a natural disaster, Cyclone Fani Date: 26 April 2019 – 5 May 2019.
- Next, we use these expressions of solidarity to characterize human behavior in online social networks, through the temporal and geospatial diffusion of emojis, exemplified by movements related to #BlackLivesMatter, #MeToo

Using Emojis to Understand Human Behavior

With respect to research on expressional solidarity, it is found that individuals were more outspoken on social media after a tragic event. They studied solidarity in tweets spanning geographical areas and several languages related to a crisis event. Extant research on emojis usage has designated three categories, that are,

1) function: when emojis replace a conjunction or prepositional word.

An example of this category would be “I 🍩 like you”, to be read as “I do not like you”

2) content: when emojis replace a noun, verb, or adjective

An example of this would be “The 🔑 to success”, to be read as “the key to success”.

3) multimodal: when emojis are used to express an attitude, the topic of the message or communicate a gesture. For example, a cool attitude can be shown as 🙌 and 😎 clapping gesture can be showed by.

Sentiment Analysis

Sentiment analysis uses data mining processes and techniques to extract and capture data for analysis in order to discern the subjective opinion of a document or collection of documents, like blog posts, reviews, news articles and social media feeds like tweets and status updates.

Sentiment analysis allows organizations to track the following:

- Brand reception and popularity
- New product perception and anticipation
- Company reputation



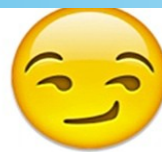
**HAPPY FACE
WITH WIDE
MOUTH AND
SQUINTING
EYES**



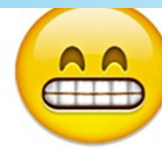
**RELIEVED
FACE**



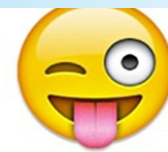
**FACE WITH
HEART
SHAPED EYES**



**SMIRKING
FACE**



**GRIMMACING
FACE**



**WINKING
FACE WITH
STUCK OUT
TONGUE**



**FACE WITH
STUCK OUT
TONGUE AND
EYES CLOSED**



**DISAPPOINTED
FACE**



**KISSING FACE
WITH CLOSED
EYES**



**HAPPY AND
CRYING FACE**



SLEEPY FACE



**FACE WITH
COLD SWEAT**



**LOUDLY
CRYING FACE**



**HAPPY FACE
WITH OPEN
MOUTH**



**HAPPY FACE
WITH OPEN
MOUTH AND
SQUINTING
EYES**



**WINKING
FACE**



**EXPRESSION-
LESS FACE**



**PERSON
FROWNING**

This is an example showing a number of emojis which have a direct relation with human sentiments.

The Emojis of Solidarity

The main goal of this article is to investigate how individuals use emojis to express a pro-social behavior, in this case, solidarity, during crisis events. Accordingly, we outline our analyses in the form of RC (Research Challenges).

RC1: How useful are emojis as features in classifying expressions of solidarity?

Standard pre-processing techniques are used for removing stopwords and lowercasing the tweets. The hashtags that were annotated from the tweets are removed and this raw cleaned data is used for processing of the result.

EXAMPLE



Priyam Mukherjee

@MeMester_9748

I am the MaN of my Words 😊😂

11:29 AM · 15 May 19 · [Twitter for Android](#)

The - Stopwords (words with least meaning).
MaN – Togglecase Letters.

Words 😊😂 - Emojis describing the state of mind.

RC2: Which emojis are used in expressions of solidarity during crisis events and how do they compare to emojis used in other tweets?

The emojis used during crisis events are very much used in large number compared to other emojis which are not much linked with crisis event.

For eg: Terrorist attack at a certain place calls for the emojis 🙏 and 😞 over other images.

So, the emojis used in expressions of solidarity is very much dependent on the crisis events.

RC3: Which emojis occur in tweets that are posted within areas directly affected by crisis events as compared to those tweets that are posted from other areas?



This research question puts forth that solidarity would be expressed differently by people that are directly affected by the crisis than those who are not.



For eg: The recent terror attacks in Sri Lanka would call for 🙏 for grief, 😡 for anger towards the terrorists, 🇱🇰 Flag for supporting the country (among the people of the country) and people who did not face the situation would call for 🙏 for praying for the victims and 🇱🇰 flag by supporting the country in tough times

RC4: How can emojis be used to study the temporal and geographical diffusion of solidarity expressions during crisis events?

Our primary aim with this research question was to look at how the emojis of solidarity diffuse over time within the affected community and compare communities not affected by the same event.

(different portrayals of emojis from different communities [one community affected and other community portraying feels regarding the other one])

For example, after an earthquake at a certain place
Emojis  depicts the tension among the people,  praying to God at the time of crisis

And emojis  depicts radioactivity if happens and depicting danger at that certain place and  depicting help.

What is Machine Learning?

Aspect of AI: creates knowledge

Definition:

“changes in [a] system that enable [it] to do the same task or tasks drawn from the same population more efficiently and more effectively the next time.”

There are two ways that a system can improve:

1. By acquiring new knowledge
2. By adapting its behavior

solving problems more accurately and efficiently.

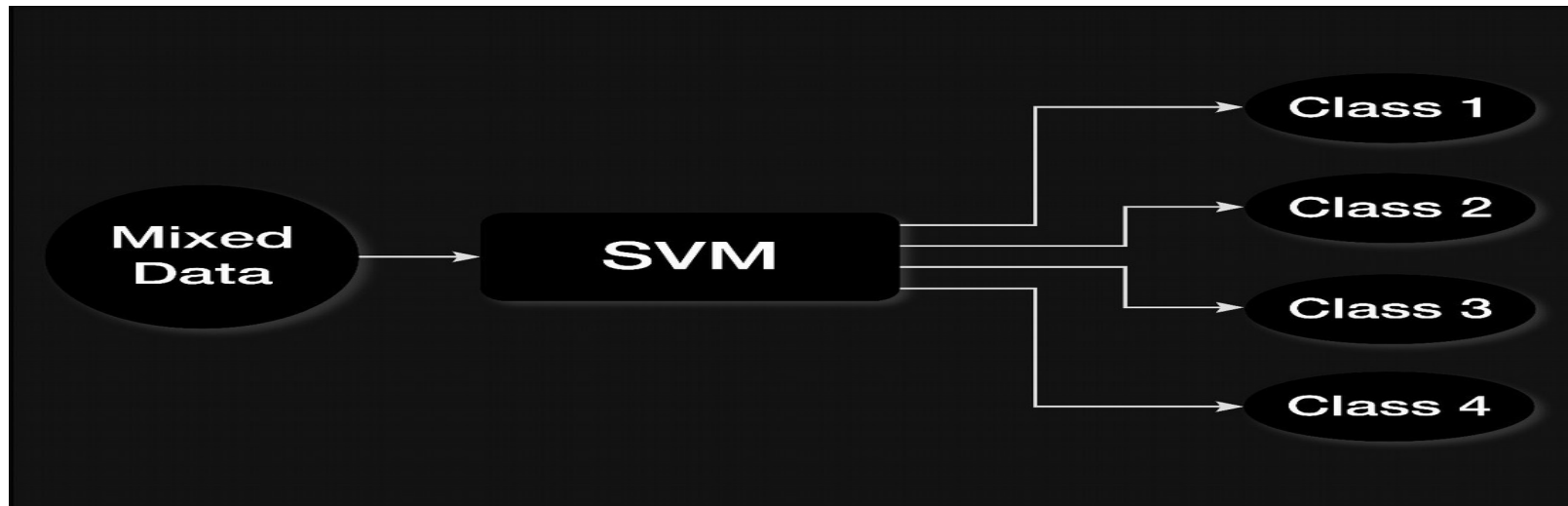
Baseline Models

We will use Support Vector Machine (SVM) with a linear kernel and 10 fold cross validation to classify tweets containing solidarity expressions. For the baseline models, we will experiment with three variants of features including (a) word bigrams, (b) TF-IDF [MPH08], (c) TF-IDF+Bigrams.





What is Support Vector Machine (SVM)?

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n -dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well.

Working of SVM



Raw unclean data is brought into the SVM and with a hyperplane the data is divided into four classes. For example, if we take Terrorist attack into account, then,

-  - will be kept in Class 1 depicting grief.
-  - will be kept in Class 2 depicting support.
-  - will be kept in Class 3 depicting anger.
-  - will be kept in Class 4 depicting nationality.

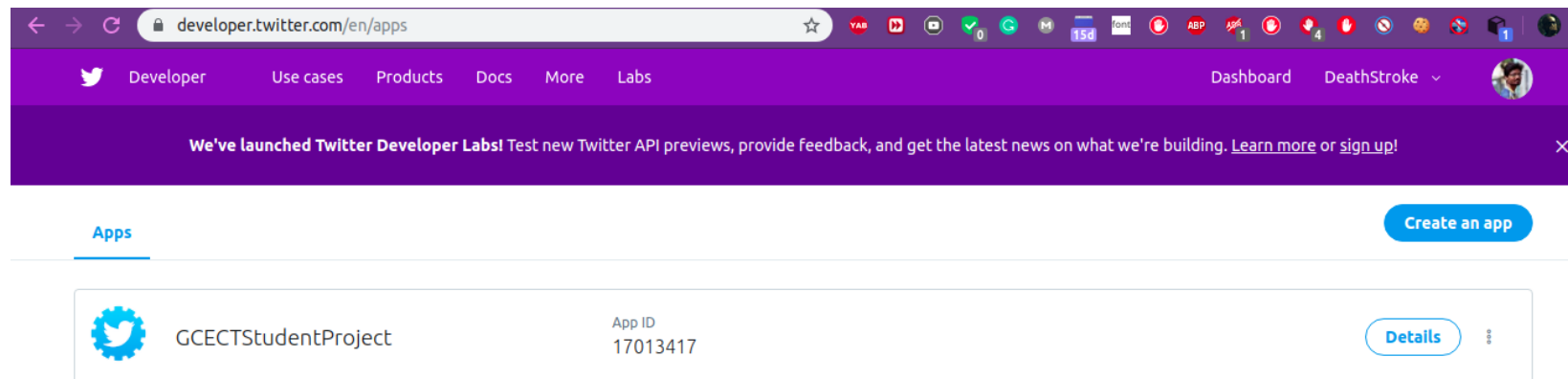
Data Collection

- Building a Twitter application using twitter api
- Using the credentials of that Twitter application (Consumer key, Consumer secret key, Access token, Access token secret) we started streaming the tweets.
- Building a simple application to show what is sentiment analysis.

Building a Twitter application using twitter api

1. Visit the Twitter Developers Site and Create a New Application.

(<https://developer.twitter.com/en/apps/17013417>)



2. Fill in the Application Details

App details

The following app details will be visible to app users and are required to generate the API keys needed to authenticate Twitter developer products.

App name (required) ⓘ

Maximum characters: **32**

Application description (required)

Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

3. Creating credentials

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys

Regenerate

Access token & access token secret

Read and write (Access level)

Revoke

Regenerate

4. Choose what Access Type You Need

Change application type to 'Read & Write' to give the application permission to follow other accounts on your behalf. This will require that you verify your mobile phone within your twitter account.

Application Type

Access:

- ☒ Read only
- ☐ Read and Write
- ☐ Read, Write and Access direct messages

What type of access does your application need? Note: @Anywhere applications require read & write access.
Find out more about our [Application Permission Model](#).

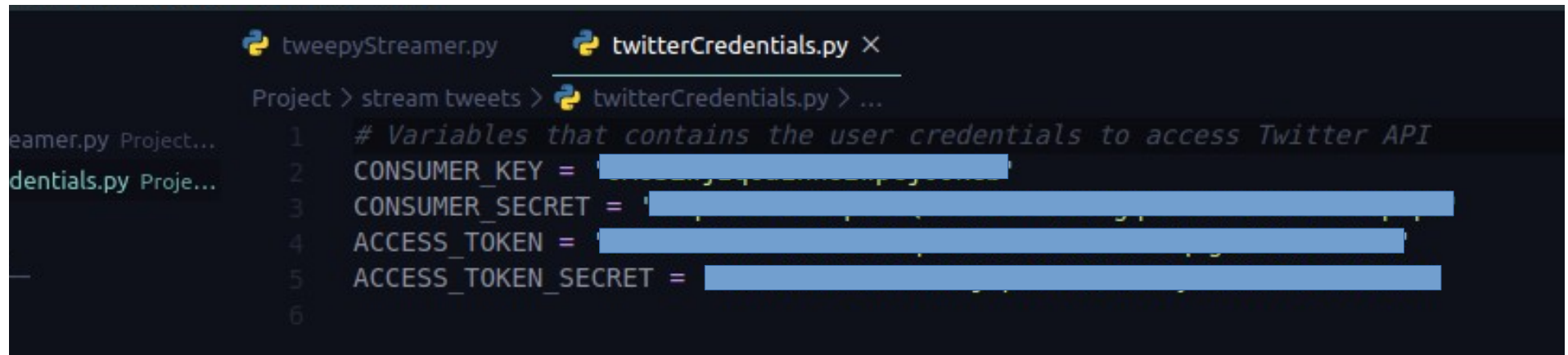
Streaming Tweets

Step 1: For user authentication

- Consumer key, Consumer secret key

Step 2: For accessing the data

- Access token, Access token secret



```
tweepyStreamer.py  twitterCredentials.py X
Project > stream tweets > twitterCredentials.py > ...
1  # Variables that contains the user credentials to access Twitter API
2  CONSUMER_KEY = '...'
3  CONSUMER_SECRET = '...'
4  ACCESS_TOKEN = '...'
5  ACCESS_TOKEN_SECRET = '...'
6
```

Step 3: Streaming tweets using a listener object

After streaming the tweets, we stored the data in a .csv file.

	You will go down in history as the man who brought the #AyodhyaCase to its logical conclusion...
127	RT @ArmyVanitha: PEACE♥♥ #BabriMasjid #AyodhyaCase #AYODHYAVERDICT #AyodhyaHearing #RamMandir #AyodhyaJudgment #hindumuslimbhaibhai #JaiS...
128	Flashback: The Journey Of A Old-Decade Ram Janmaboomi-Babri Masjid Case Read Here: https://t.co/Y8TKoVXHNA... https://t.co/RmFwrSWaM0
129	RT @VamsiChandReddy: All #Congress leaders shall speak only the official stand of the @INCIndia party on the #AYODHYAVERDICT #AyodhyaCase #...
130	RT @archu243: Take a bow CJI, #RanjanGogoi You will go down in history as the man who brought the #AyodhyaCase to its logical conclusion...
131	RT @archu243: Take a bow CJI, #RanjanGogoi You will go down in history as the man who brought the #AyodhyaCase to its logical conclusion...
	RT @Deewana_Dev_: This is true India and an actual definition of peace attained after the verdict !!
132	Spread Love and Peace ! ♥

Model Building

- Importing Libraries
- Variable Declaration
- Recognizing Emojis And Setting Sentiment Ranking
- Data Pre-Processing: Removing stopwords,
- Storing hashtags, Emojis and computing sentiment scores
- Building Clusters
- Storing Scores: Final centroids, Accuracy for individual emotions, Average accuracy score

Importing Libraries

Its large and robust standard library makes Python score over other programming languages. The standard library allows you to choose from a wide range of modules according to your precise needs. Each module further enables you to add functionality to the Python application without writing additional code.

```
3 import sys
4 reload(sys)
5 sys.setdefaultencoding('utf-8')
6
7 #time for execution
8 import timeit
9 start_time = timeit.default_timer()
10
11 from nltk.corpus import sentiwordnet as sw
12 import random
13 from sklearn import svm
14 import nltk
15 import re
16 from emoji import UNICODE_EMOJI
17 from nltk.stem.snowball import SnowballStemmer
18 import os
19 import re
20 import math
21 import numpy as np
22
```

Variable Declaration

- **posWords** – This is to store the words which are responsible for positive solidarity.
- **negWords** - This is to store the words which are responsible for negative solidarity.

```
33
34 #----- nltk variables -----
35
36 words = list(set(w.lower() for w in nltk.corpus.words.words()))
37
38 stopWords = list(set(w.lower() for w in nltk.corpus.stopwords.words()))
39
40
41 stemmer = SnowballStemmer("english", ignore_stopwords=True)
42
43
44 posTweets = []
45 negTweets = []
46
47 posTweetIDs = []
48 negTweetIDs = []
49
50 #----- variable declaration -----
51
52 posWords = []
53 negWords = []
54
55
56 finalClustersDict = {1:"", 2:"", 3:"", 4:"", 5:"", 6:""}
57 finalClusters = []
58
```

Recognizing Emojis And Setting Sentiment Ranking

```
67 #-----recognize emojis-----
68
69
70 def is_emoji(s):
71     count = 0
72     for emoji in UNICODE_EMOJI:
73         count += s.count(emoji)
74         if count > 1:
75             return False
76     return bool(count)
77
78
79 # ----- emoji sentiment rank from http://kt.ijs.si/data/Emoji_sentiment_ranking/ -----
80
81 emoji_SentimentScores = {}
82
83 #happy, angry, love, sad, playful, confused
84 emoji_SentimentScores["\xF0\x9F\x98\x82"] = 0.221 #0.221*2
85 emoji_SentimentScores["\xF0\x9F\x98\xA1"] = -0.173 #-0.173
86 emoji_SentimentScores["\xe2\x9d\xa4"] = 0.746 #0.746*2
87 emoji_SentimentScores["\xF0\x9F\x98\xAD"] = -0.093 #-0.093*2
88 emoji_SentimentScores["\xF0\x9F\x98\x9C"] = 0.445 #0.445*2
89 emoji_SentimentScores["\xf0\x9f\x98\x95"] = -0.397 #-0.397*2
```

- **is_emojis** – To recognize the emoticons from the tweets we have used this function.
- **emoji_SentimentScores** – After finding the emojis we have calculated and stored the **Sentiment Score** for each emoticons into this python dictionary. We have taken **Sentiment Ranking** of emoticons from a third party website given below:

https://ktijs.si/data/Emoji_sentiment_ranking/

Data Pre-Processing: Removing stopwords, storing hashtags, emojis and computing sentiment scores

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

```
107 #happy, angry, love, sad, playful, confused
108 for row in unclean_file.readlines():
109     #remove usernames
110     row = ' '.join(re.sub("([@A-Za-z0-9_]+)", "", row).split())
111
112     #remove stopWords
113     wordList = row.split()
114     for word in wordList:
115         if word in stopWords or len(word) == 1:
116             row = row.replace(word, "")
117             #print (word, row)
118     try:
119         tweets.append(row)
120     except:
121         pass
122
123 #happy, love, playful, sad, angry, confused
124 targets = [0]*len(tweets)
125
126 for target in range(len(tweets)):
127     if("\xF0\x9F\x98\x82" in tweets[target]):
```

Building Clusters

Clustering describes the task of discovering groups of similar items with a large collection. It is also describe as unsupervised machine learning, as the data from which it learns is unannotated with class information, as is the case for supervised learning.

```
514
515 finalClustersDict = {1:"", 2:"", 3:"", 4:"", 5:"", 6:""}
516 finalClustersIdx = []
517
518 k = 0
519 for clusters in finalClusters:
520     finalClustersIdx.append([])
521
522     for cluster in clusters:
523         myidx = finalCentroids.index(min(finalCentroids, key=lambda x: abs(x - cluster)))+1
524         #print cluster, centroids, myidx
525         finalClustersIdx[k].append(myidx)
526     k += 1
527
528 indices = []
529 for cluster in finalClustersIdx:
530     indices.append(max(cluster,key=cluster.count))
```

Storing Scores: Final Centroids, Accuracy for individual emotions, Average accuracy score

```
527
528 indices = []
529 for cluster in finalClustersIdx:
530     indices.append(max(cluster, key=cluster.count))
531
532 #print indices
533 accuracy = [0]*len(finalCentroids)
534
535 k = 0
536 for cluster in finalClustersIdx:
537     for row in cluster:
538         if(row == indices[k]):
539             accuracy[k] += 1
540         k += 1
541
542 for row in range(len(accuracy)):
543     #print accuracy[row], len((finalClusters[row]))
544     accuracy[row] = accuracy[row]/float(len(finalClusters[row]))*100
545
546 print "Accuracy for individual emotions", accuracy
547 print "Average accuracy", sum(accuracy)/len(accuracy)
548
```

Centroids			
Emoticon	Our-code	skfuzzy	kmeans
happy	0.2014	0.2105	0.4504
playful	0.4706	0.4515	0.7829
love	0.9443	0.7732	2.0397
sad	-0.1191	-0.0806	0.1922
angry	-0.3858	-0.1884	-0.1393
confused	-0.6466	-0.4485	-0.4719

Cluster Centroid: The middle of a cluster. A centroid is a vector that contains one number for each variable, where each number is the mean of a variable for the observations in that cluster. The centroid can be thought of as the multi-dimensional average of the cluster.

Results

```
ujaan@Ujaan: ~/x/Project
File Edit View Search Terminal Help
(base) ujaan@Ujaan:~/x/Project$ conda activate python2.7
(python2.7) ujaan@Ujaan:~/x/Project$ python sentimentAnalysis.py
-----tweets-----
Final Centroids: [ 0.24966666  0.77713281  0.48918435 -0.12483333 -0.49933333 -0.24966667]
Accuracy for Individual Emotions: [100.0, 75.0, 100.0, 100.0, 100.0, 100.0]
Average Accuracy: 95.8333333333
('Time elapsed is ', 11.278732061386108)
(python2.7) ujaan@Ujaan:~/x/Project$
```

→ Final Centroids:

[0.24966666, 0.77713281, 0.48918435, -0.12483333, -0.49933333, -0.24966667]

→ Accuracy for Individual Emotions:

[100.0, 75.0, 100.0, 100.0, 100.0, 100.0]

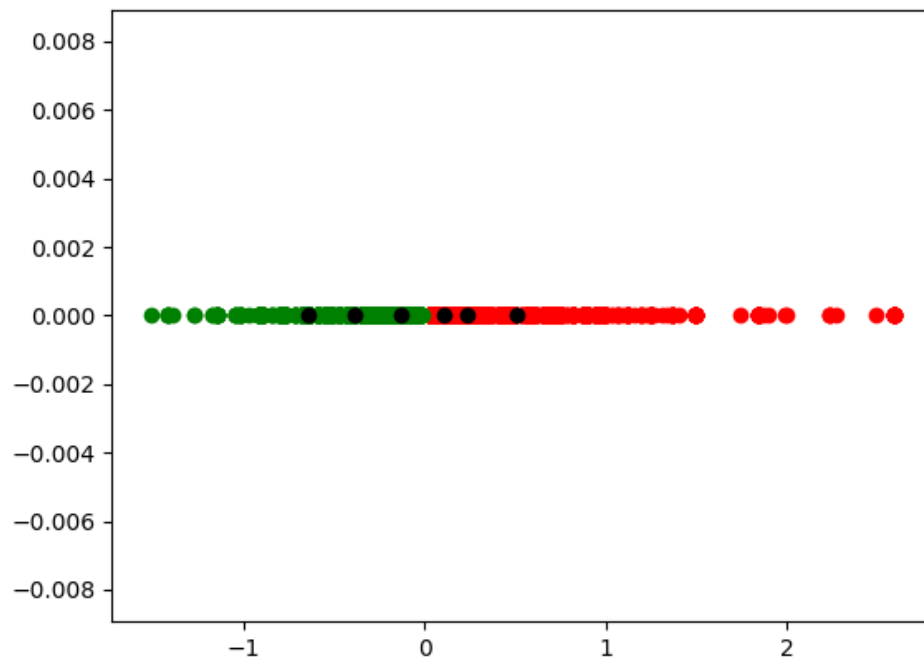
→ Average Accuracy: 95.8333333333

Data Visualization

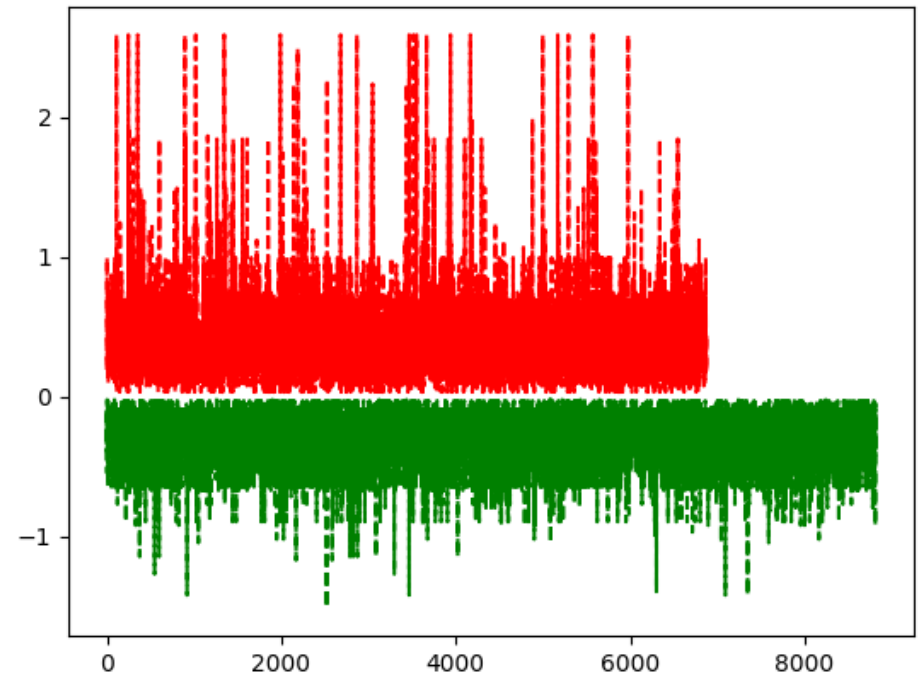
Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Importance of Data Visualization:

- We need data visualization because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet.
- Data visualization gives us a usefull insight of the data, so it is very much understandable for the human mind and brain.



1. Final Centroids:



2. Positive Tweets vs Negative Tweets:

Above figure shows the distribution of positive and negative tweets, which gives a better idea about the variance between the tweet sentiment scores. For figures 1, 2 the black dots from left to right represent the centroids for each emoticon in the order – confused, angry, sad, happy, playful and love. The figures show the variation in the final centroids obtained by using the corresponding technique. We see that our SVM works almost as well as skfuzzy FCM, while K-means doesn't perform as well as the other 2 methods. The accuracy table shows the maximum frequency of accuracy over 50 runs of each code.

Future Scope

1. Ambiguity of text and emojis can be lowered if more possibilities are taken into consideration.
2. We will use the output data in various type of Machine Learning models for various type of analysis like Text Analysis, Sentiment Analysis, Solidarity Analysis.

Conclusion

Our analysis will reveal that emojis are a powerful indicator of sociolinguistic behaviors (solidarity) that are exhibited on social media as the crisis events unfold.

An additional future goal is to analyze the interaction of sentiment of emojis and solidarity as well as the text that occurs with these emojis in further detail.

References

1. <https://www.analyticsvidhya.com/>
2. <https://www.kaggle.com/>
3. <http://www.datasciencecentral.com/>
4. <https://towardsdatascience.com/>
5. Bac Le and Huy Nguyen, Twitter Sentiment Analysis Using Machine Learning Techniques, Advances in Intelligent Systems and Computing book series, Volume 358, 2015, 978-3-319-17996-4
6. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband, Machine Learning-Based Sentiment Analysis for Twitter Accounts, Mathematical Computational Application, Volume 23, 27 February 2018, <https://doi.org/10.3390/mca23010011>

THANK YOU