International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Using Social Media Trends to Provide Disaster Early Warning Systems and Disaster Assessment

Ch. Chakradhara Rao, B. Meena, K. Pavan Kumar

Abstract: ***Tremors, floods, dry season, and other normal perils cause billions of dollars in monetary misfortunes every year around the globe. A huge number of dollars in philanthropic help, crisis credits, and advancement help are consumed every year. However endeavors to lessen the dangers of normal perils remain generally ungraceful crosswise over various risk types and don't really concentrate on regions at most astounding danger of debacle. Informal communities are assuming an undeniably significant job as early cautioning frameworks, supporting with quick debacle appraisal and post-fiasco recuperation. There is a requirement for both the general population and fiasco help offices to all the more likely see how web based life can be used to survey and react to catastrophic events. This work directs a various leveled multistage investigation dependent on numerous information assets, consolidating internet based life information and monetary misfortunes. This work attracts regard for the way that during a catastrophe, residents go to internet based life and most of tweets contain data about the tropical storm as well as its contact with negative estimation. This paper researches whether the mix of web based life and geo-area data can add to an increasingly proficient early cautioning framework and help with calamity evaluation.***

Keywords: Social media, Geo-location information, Disaster early warning, Disaster assessment, cyclones.

I. INTRODUCTION

Slant investigation an approach to assess composed or communicated in language to decide whether the articulation is good, negative, or nonpartisan, and to what degree. The present calculation based slant investigation apparatuses can deal with gigantic volumes of client criticism reliably and precisely. Combined with content examination, supposition investigation uncovers the client's assessment about themes running from your items and administrations to your area, your ads, or even your rivals. It alludes to the utilization of normal language preparing, content examination, computational semantics and biometrics to efficiently distinguish, remove, measure and concentrate full of feeling states and emotional data.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Ch. Chakradhara Rao*, CSE department, Raghu Engineering College, Visakhapatnam, India. Email: chakri.rec@gmail.com
B. Meena, CSE department, Raghu Engineering College, Visakhapatnam, India. Email: meenacserec@gmail.com
K. Pavan Kumar, CSE department, Raghu Engineering College, Visakhapatnam, India. Email:pavankumar.krishnardhula@gmail.com

Notion investigation is broadly applied to voice of the client materials, for example, audits and overview reactions, on the web and online life, and social insurance materials, for application that range from showcasing to client care to clinical drug.
These are of following types:

Text sentiment analysis

Numerous specialists have formulated different calculations to dissect printed material for foreseeing opinion. A recursive neural system classifier alluded to as CoreNLP. CoreNLP depends on the common language pipeline that contains a few stages including tokenization (i.e., hacking content into words, numbers, spaces, and accentuations), grammatical form tagger (i.e., doling out a syntactic classification to every token), and morphological examination (i.e., finding the root type of each word). Another notable estimation classifier is NLTK, the Natural Language Toolkit for Python, which uses the Porter stemming calculation (a streamlined strategy for morphological examination) and depends on the credulous Bayes classifier. Moreover, SentiStrength which concentrates the power of content by assessing the quality of positive and negative of every token in the content. SentiStrength indicated great outcomes for social web writings (i.e., short message). Then again, a few business content opinion classifiers are as of late offered by the business (e.g., Microsoft Azure Text Analytics API, Google Cloud Natural Language API, and IBM Watson Alchemy Language API).

Virtual sentiment analysis

As of late, specialists have researched separating assumption from visual substance on the grounds that visual information have turned into a basic piece of web based life. Current methodologies of opinion investigation on pictures are sorted into three components: low-level visual element based techniques, mid-level visual element based strategies, and profound learning based methodologies (i.e., the convolutional neural system). Utilized shading highlights for the visual assessment forecast and proposed a visual opinion metaphysics (alluded to as SentiBank) made out of 1200 descriptive word thing sets. Like SentiBank, Yuan et al. utilized 102 scene traits. Among the examination endeavors which used the convolutional neural system (CNN), Campos et al. applied exchange figuring out how to tweak the neural system with Flickr Dataset. In the business, there are business items for outward appearance expectation (e.g., Microsoft Azure Computer Vision API and Google Cloud Vision API) yet these items are not ready to foresee the notion of "free of faces" pictures.

On the other hand, these items offer different capacities for extricating watchwords portraying picture's substance. Therefore, content assessment classifier can be utilized on the blend of the separated watchwords. All in all, both content and visual feeling classifiers produce a notion mark (i.e., s) as negative, nonpartisan, or positive which relates to the numeric assumption score - 1, 0, and 1, individually.

Why is sentiment analysis important?

Feeling investigation is basic since encourages you see what clients like and aversion about you and your image. Client criticism from online networking, your site, your call focus operators, or some other source contains a fortune trove of helpful business data. Be that as it may, it isn't sufficient to realize what clients are discussing. You should likewise know how they feel. Notion investigation is one approach to reveal those emotions. Now and again known as "supposition mining," assessment examination can fill you in as to whether there has been an adjustment in general feeling toward any part of your business. Pinnacles or valleys in slant scores give you a spot to begin on the off chance that you need to make item upgrades, train deals or client care operators, or make new promoting efforts

II. PROPOSED SYSTEM:

In this paper, we build up a structure which consolidates information from various sources, including web based life (Twitter) and geo-area (in the case of being waterfront and the vicinity to the storm focus), with data on calamity misfortunes. To start with, we study how natives respond to a catastrophe utilizing investigations of both web based life volume and substance.

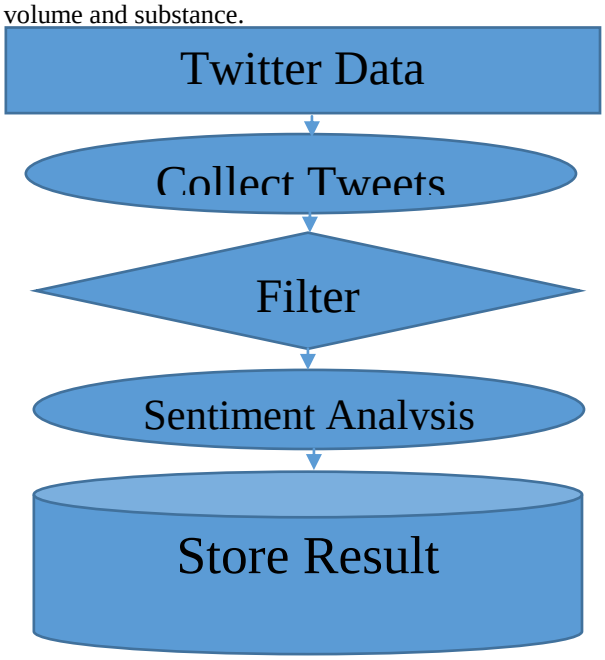


Fig. 1. Usecase diagram

This considers the investigation of whether the mix of online life and geo-information has a preferred position in early cautioning. At last, we consolidate data from catastrophe related Twitter movement and geo-area highlights to dissect their relationships with the harm brought about by Hurricane Harvey.

Related Work:

The client who needs to utilize this task ought to have the learning about the different datasets and information mining ideas particularly unique regular itemset mining procedures and brief thought on the most proficient method to locate the unpromising things from the datasets r.

Explanation:

There are two on-screen characters present in this usecase outline Fig. 1. They are "client" and "framework". Client recovers the exchanges from the dataset given by the framework. Here, the client can choose a calculation and execute that calculation by passing required information esteems.

Naïve Bayes Classifier:

A guileless Bayes classifier utilizes likelihood hypothesis to group information. Credulous Bayes classifier calculations utilize Bayes' hypothesis. The key knowledge of Bayes' hypothesis is that the likelihood of an occasion can be balanced as new information is presented.

What makes a credulous Bayes classifier gullible is its supposition that all traits of an information point under thought are free of one another. A classifier arranging natural products into apples and oranges would realize that apples are red, round and are a sure size, however would not accept every one of these things on the double. Oranges are round as well, all things considered.

An innocent Bayes classifier is certifiably not a solitary calculation, however a group of AI calculations that make employments of measurable freedom. These calculations are generally simple to compose and run more effectively than progressively complex Bayes calculations.

The most prevalent application is spam channels. A spam channel sees email messages for certain catchphrases and places them in a spam envelope in the event that they coordinate.

Regardless of the name, the more information it gets, the more exact a gullible Bayes classifier turns out to be, for example, from a client hailing email messages in an inbox for spam.

NLTK:

The Natural Language Toolkit (NLTK) is a stage utilized for structure Python programs that work with human language information for applying in measurable regular language preparing (NLP).

It contains content handling libraries for tokenization, parsing, order, stemming, labeling and semantic thinking. It likewise incorporates graphical showings and test informational indexes just as joined by a cook book and a book which clarifies the standards behind the fundamental language preparing undertakings that NLTK bolsters.

The Natural Language Toolkit is an open source library for the Python programming language initially composed by Steven Bird, Edward Loper and Ewan Klein for use being developed and training.

It accompanies a hands-on guide that presents subjects in computational semantics just as programming essentials for Python which makes it appropriate for language specialists who have no profound information in programming, designers and scientists that need to dive into computational phonetics, understudies and teachers.

NLTK incorporates in excess of 50 corpora and lexical sources, for example, the Penn Treebank Corpus, Open Multilingual WordNet, Problem Report Corpus, and Lin's Dependency Thesaurus.

PANDAS:

PANDAS is an open source, BSD-authorized library giving superior, simple to-utilize information structures and information examination instruments for the Python programming language.

It is a NumFOCUS supported task. This will help guarantee the achievement of advancement of pandas as a world-class open-source venture and makes it conceivable to give to the undertaking.

Java is an Object-Oriented Programming language created by Sun Microsystems. In spite of the fact that online projects are quality of the language, java is universally useful language that can be utilized to build up a wide range of projects. It is accumulated to at least one records of byte codes with the expansion class.

PANDAS don't execute critical displaying usefulness outside of straight and board relapse; for this, look to details models and scikit-learn. More work is as yet expected to make Python a top notch factual demonstrating condition, however we are well on our way toward that objective.

NUMPY:

NumPy is a universally useful cluster preparing bundle. It gives an elite multidimensional cluster article, and instruments for working with these exhibits.

It is the crucial bundle for logical processing with Python. It contains different highlights including these significant ones:

- A ground-breaking N-dimensional exhibit object
- Sophisticated (telecom) capacities
- Tools for coordinating C/C++ and Fortran code
- Useful direct variable based math, Fourier change, and arbitrary number capacities

Other than its conspicuous logical uses, NumPy can likewise be utilized as an effective multi-dimensional compartment of nonexclusive information.

Matplot lib:

Matplotlib is a Python 2D plotting library which produces distribution quality figures in an assortment of printed copy designs and intuitive situations crosswise over stages. Matplotlib can be utilized in Python contents, the Python and IPython shells, the Jupyter note pad, web application servers, and four graphical UI toolboxs.

Methods:

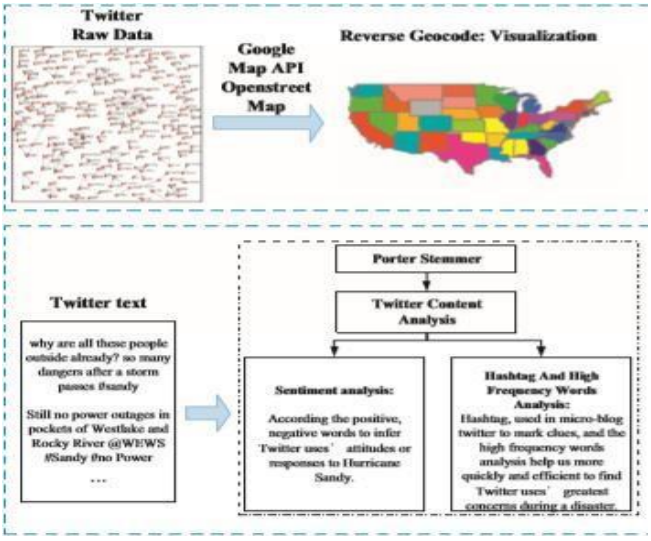


Fig. 2. Block diagram

Data Preprocessing:

Data pre-processing is an information mining system that includes changing crude information into a justifiable organization. True information is frequently deficient, conflicting, as well as ailing in specific practices or drifts and is probably going to contain numerous mistakes. It is a demonstrated strategy for settling such issues. Information pre-handling gets ready crude information for further preparing. On the off chance that there is much unessential and repetitive data present or uproarious and temperamental information, at that point learning revelation during the preparation stage is progressively troublesome. Information arrangement and separating steps can take significant measure of preparing time. It incorporates cleaning, occasion determination, standardization, change, highlight extraction and choice.

Data Cleaning:

Data cleansing or data cleaning is the way toward recognizing and remedying (or expelling) degenerate or mistaken records from a record set, table, or database and alludes to distinguishing inadequate, erroneous, incorrect or superfluous pieces of the information and after that supplanting, adjusting, or erasing the filthy or coarse information. It might be performed intelligently with information wrangling devices, or as bunch handling through scripting. In the wake of purging, an informational collection ought to be steady with other comparable informational collections in the framework. The irregularities distinguished or expelled may have been initially brought about by client section blunders, by debasement in transmission or capacity, or by various information word reference meanings of comparable elements in various stores. Data cleaning varies from information approval in that approval constantly implies information is rejected from the framework at passage and is performed at the hour of section, as opposed to on clusters of information.

Data Integration:

Data integration includes consolidating information living in various sources and furnishing clients with a brought together perspective on them. This procedure winds up critical in an assortment of circumstances, which incorporate both business, (for example, when two comparable organizations need to consolidate their databases) and logical (joining exploration results from various bioinformatics storehouses, for instance) spaces. Data integration shows up with expanding recurrence as the volume (that is, huge information) and the need to share existing information detonates. It has turned into the focal point of broad hypothetical work, and various open issues stay unsolved.

Data Transformation:

Data transformation is the way toward changing over information or data starting with one arrangement then onto the next, more often than not from the configuration of a source framework into the required organization of another goal framework. The typical procedure includes changing over archives, yet information transformations some of the time include the transformation of a program starting with one coding languages then onto the next to empower the program to keep running on an alternate stage.

The typical purpose behind this information movement is the appropriation of another framework that is entirely unexpected from the past one.

Sentiment Analysis:

Sentiment analysis is an approach to assess composed or communicated in language to decide whether the articulation is ideal, horrible, or unbiased, and to what degree. The present calculation based assessment examination apparatuses can deal with colossal volumes of client criticism reliably and precisely. Matched with content examination, feeling investigation uncovers the client's assessment about subjects going from your items and administrations to your area, your ads, or even your rivals. It alludes to the utilization of common language preparing, content investigation, computational semantics and biometrics to deliberately distinguish, remove, evaluate and study emotional states and abstract data. Assumption investigation is broadly applied to voice of the client materials, for example, audits and study reactions, on the web and online networking, and social insurance materials, for application that range from promoting to client care to clinical medication.

Reverse Geocoding:

To encourage spatial investigation of tweets, we additionally built up a geotagging module that shows the substance of a tweet at its geographic area on a guide. We do this by utilizing a tweet's directions if it's geotagged, or the area data from the client's profile. In particular, if a tweet is geotagged, we show it at its scope/longitude arranges. Else, we utilize the area field of the client profile to decide a scope/longitude position. We first pass the area string to the Yahoo geocoding administration and recover the best five matches around the world. We at that point select the most appropriate one utilizing state or nation imperatives. Fig. 2 demonstrates a case of geotagging tweets from the February 2011 earth-shake. This figure shows the appropriation of tweets that can be geo-labeled; the marker hues demonstrate the volume of tweets caught at a particular area.

III. RESULTS AND DISCUSSIONS

Twitter data

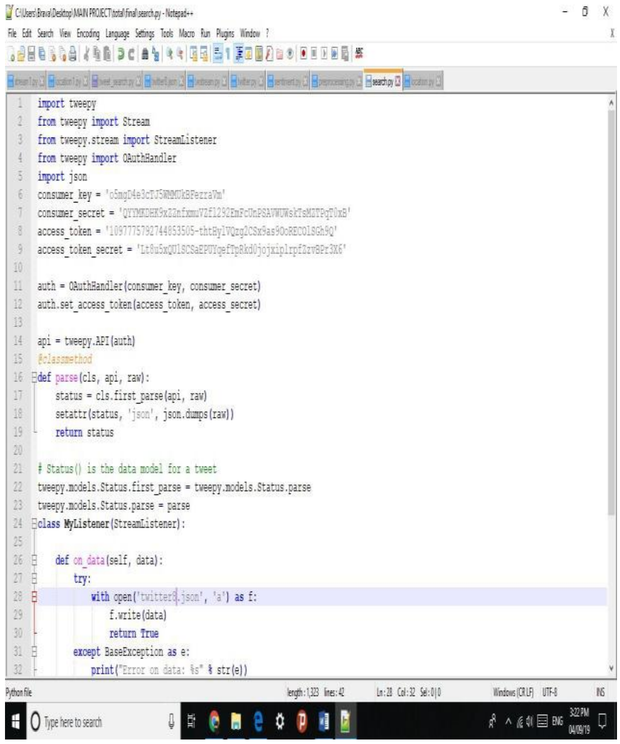


Fig. 3. Data collection code

Fig. 3 shows the code to create an API in the twitter developer's website and generate consumer keys and access token keys. These keys will be helpful to collect the data from the twitter. In this code the hurricane Harvey taken as a search term and related tweets were collected based on hashtag in a json file.

The main advantage of this code is to retrieve the millions of tweets which contains search term.

Collection of Tweets:

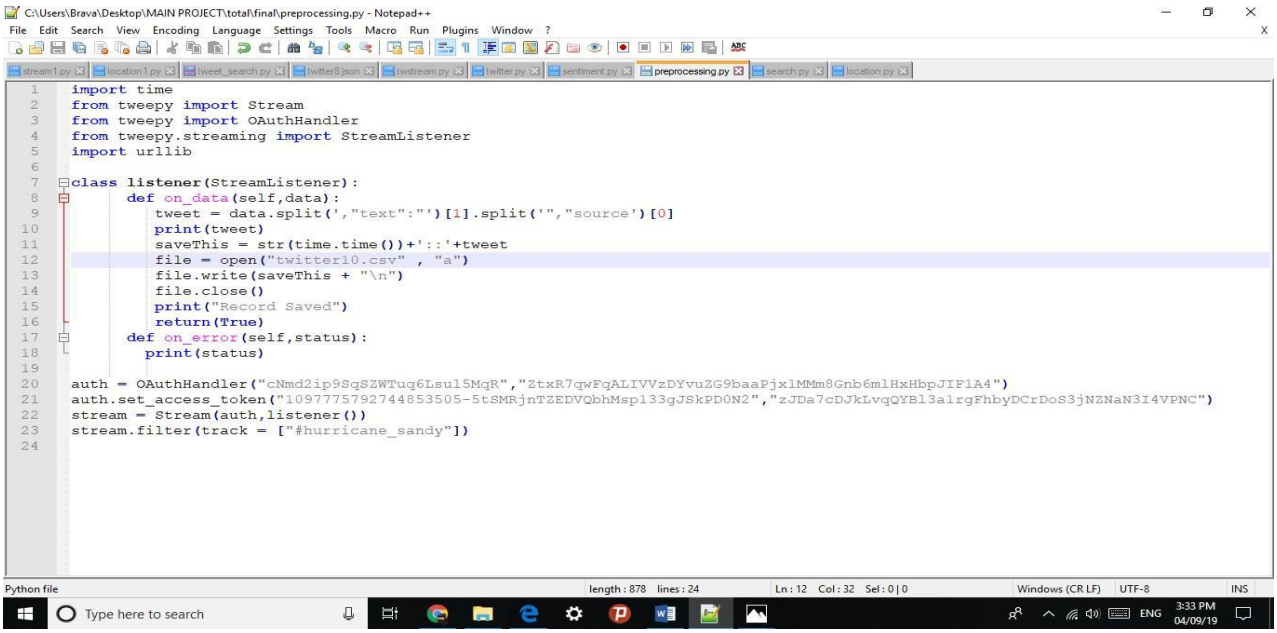


Fig. 4. Data preprocessing

In this, processing takes place. Data preprocessing is a proven method of resolving such issues. Data cleaning is process of fill in missing values, smoothing the noisy data, identify or remove outliers, and resolve inconsistencies. Integration of multiple databases, data cubes, or files.

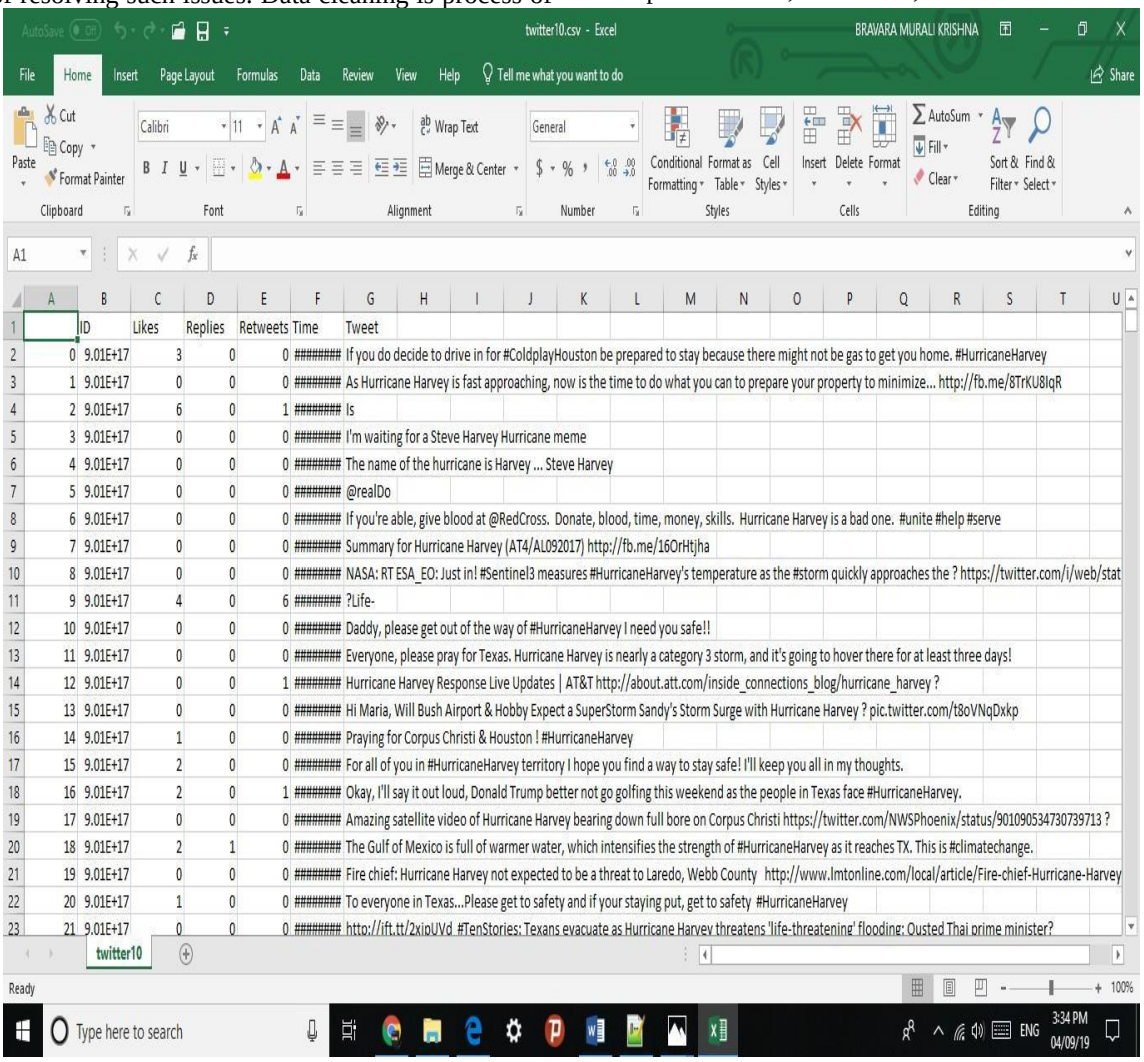
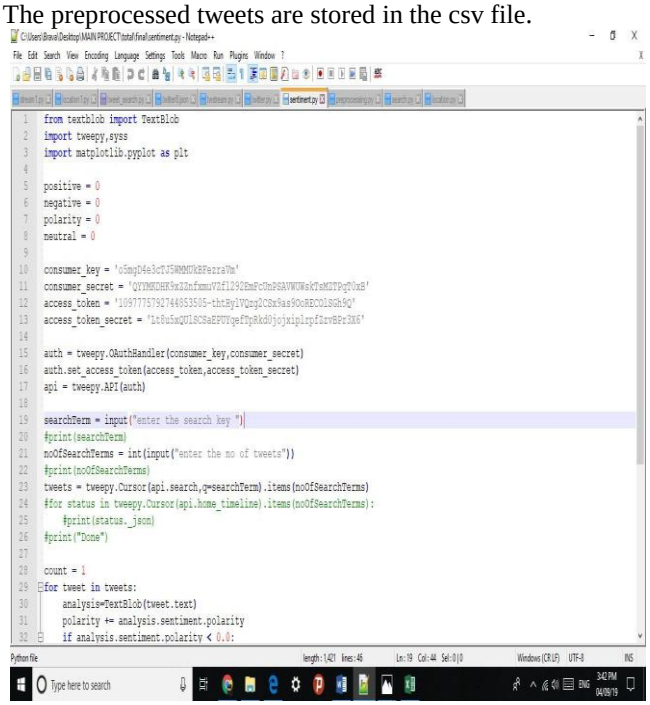


Fig. 5. Preprocessed data

The preprocessed tweets are stored in the csv file.



```
from textblob import TextBlob
import tweepy,sys
import matplotlib.pyplot as plt

positive = 0
negative = 0
polarity = 0
neutral = 0

consumer_key = 'o5mgD4e3cTJ5WMMTkBPezraYm'
consumer_secret = 'QYYMKDHK5w33nfxmuY2fl292EmFcDnPSAVWUWsKTsMZTPqYUxB'
access_token = '1097775792744853505-thtRylYQzg2CSx9as9OoREC0l8Gk9q'
access_token_secret = 'It8u5wQDl8C8aEPOYqefYpBkdOjojxiplrpfIzvBPr3R6'

auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_token_secret)
api = tweepy.API(auth)

searchTerm = input("enter the search key ")
#print(searchTerm)
noOfSearchTerms = int(input("enter the no of tweets"))
#print(noOfSearchTerms)
tweets = tweepy.Cursor(api.search,q=searchTerm).items(noOfSearchTerms)
#for status in tweepy.Cursor(api.home_timeline).items(noOfSearchTerms):
    #print(status._json)
#print("Done")

count = 1
for tweet in tweets:
    analysis=TextBlob(tweet.text)
    polarity += analysis.sentiment.polarity
    if analysis.sentiment.polarity < 0.0:
```

Fig. 6. Sentiment analysis

Fig 6 shows the code for sentiment analysis to find positive, negative, neutral of the tweets to find polarity. Twitter sentiment analysis has attracted significant public and research interest for several years, especially regarding its applications to the various stock markets, avenues of public opinion monitoring, and disaster warnings. Patterns of emotional changes revealed in social media postings are thought to reflect the spatiotemporal mood variations of society. By analyzing the tweets posted during an event or disaster.

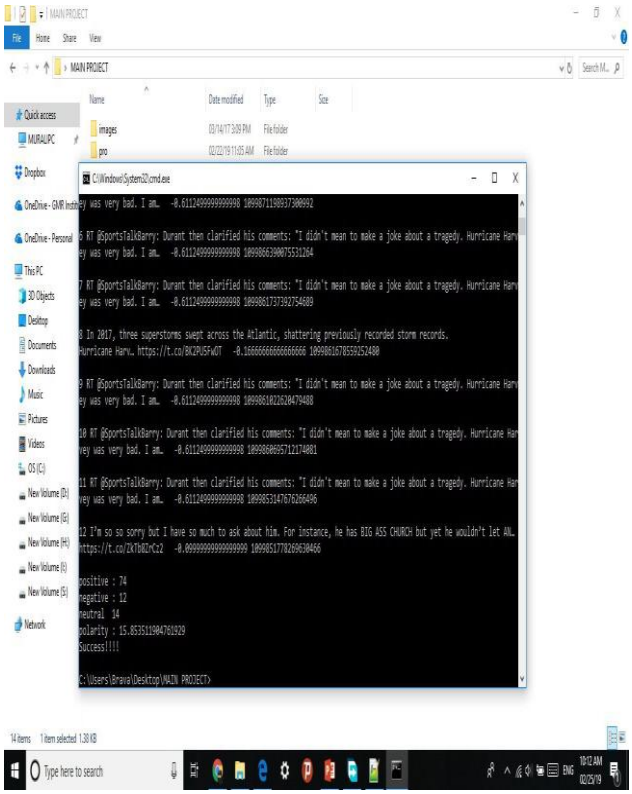
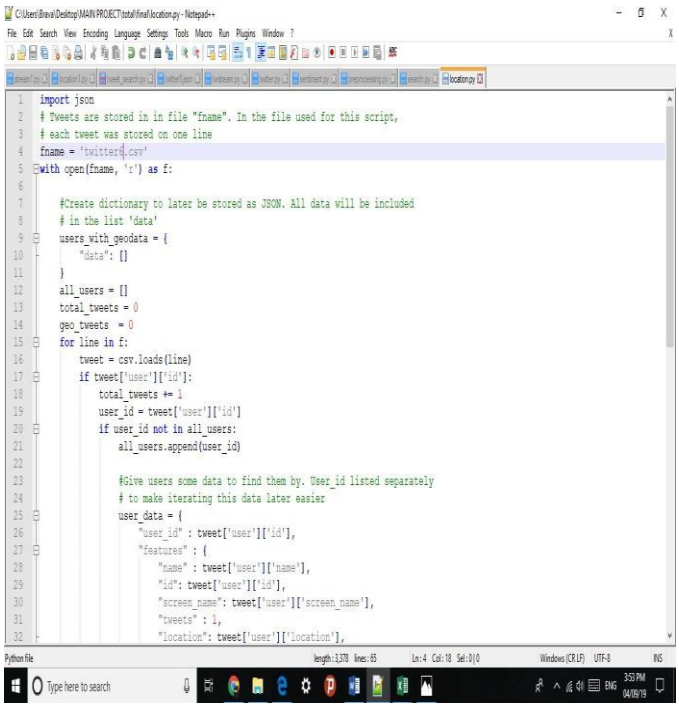


Fig. 7. Sentiment analysis output



```
import json
# Tweets are stored in in file "fname". In the file used for this script,
# each tweet was stored on one line
fname = 'twitter4.csv'
with open(fname, 'r') as f:

    #Create dictionary to later be stored as JSON. All data will be included
    # in the list 'data'
    users_with_geodata = {
        "data": []
    }
    all_users = []
    total_tweets = 0
    geo_tweets = 0
    for line in f:
        tweet = csv.loads(line)
        if tweet['user']['id']:
            total_tweets += 1
            user_id = tweet['user']['id']
            if user_id not in all_users:
                all_users.append(user_id)

                #Give users some data to find them by. User_id listed separately
                # to make iterating this data later easier
                user_data = {
                    "user_id" : tweet['user']['id'],
                    "features" : {
                        "name" : tweet['user']['name'],
                        "id": tweet['user']['id'],
                        "screen_name": tweet['user']['screen_name'],
                        "tweets" : 1,
                        "location": tweet['user']['location'],
```

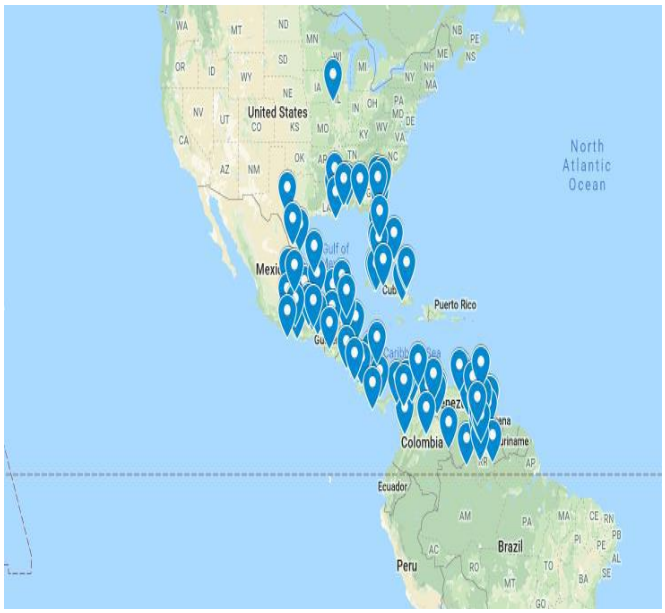


Fig. 8. PLOTTING GEO LOCATIONS

IV. CONCLUSION

Society will in general go to online life, which can reflect human life examples and rhythms that are intently connected with the spatiotemporal circulations of fiasco. Investment of government organizations and the connections of interchanges authorities with natives by means of web based life stages fortify the significance of internet based life as a marker of open familiarity with a catastrophe in neighborhoods, is exceptionally applicable to the impact. In light of the association between cataclysmic events and web based life, we lead a multi-dimensional investigation to investigate their particular connections.

In this task an information preparing system has been created dependent on switch geocoding, slant examination, hashtag and high-recurrence ways to deal with arrange the multisource information and direct a multidimensional investigation.

REFERENCES

1. R. Mechler, Natural Disaster Risk Management and Financing Disaster Losses in Developing Countries, vol. 1, Verlag Versicherungswirtsch, 2004.
2. M. Dilley, Natural Disaster Hotspots: A Global Risk Analysis, vol. 5, World Bank Publications, 2005.
3. S. Cole,H. A, E. Werker, Dovotersdemand responsive governments? Evidence from Indian disaster relief, Journal of Development Economics 97 (2) (2012) 167–181
4. R.J. Haarsma, W. Hazeleger, C. Severijns, H. Vries, A. Sterl, R. Bintanja, ... H.W. Brink, More hurricanes to hit western Europe due to global warming, Geophysical Research Letters 40 (9) (2013) 1783–1788.
5. J. Tollefson, Hurricane Sandy Spins up Climate Discussion, Nature News, 2012 (doi, 10).
6. M. Erdik, K. Şeşetyan, M. Demircioğlu, U. Hancılar, C. Zülfikar, Rapid earthquake loss assessment after damaging earthquakes, Soil Dynamics and Earthquake Engineering 31 (2) (2011) 247–266.
7. Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, M.Cebrian, Rapid assessment of
8. disaster damage using social media activity, Science Advances 2 (3) (2016) e1500779.
9. P. Bergeijk, S. Lazzaroni, Macroeconomics of natural disasters: strengths and weaknesses of meta-analysis versus review of literature, Risk Analysis 35 (6) (2015) 1050–1072.
10. S. Hallegatte, Modeling the role of inventories and heterogeneity in the assessment of the economic costs of natural disasters, Risk Analysis 34 (2014) 152–167.
11. L.N.Rickard,Z.J.Yang,J.P.Schuldt,G.M.Eosco,C.W.Scherer,R.A.Dazi a no,Sizing up a superstorm: exploring the role of recalled experience and attribution of responsibility in judgments of future hurricane risk, Risk Analysis 37 (12) (2017) 2334–2349.
12. B. Nelson, Natural disasters: a calculated risk, Nature 495 (7440) (2013) 271– 273.

AUTHORS PROFILE

Ch. Chakradhara Rao, Assistant Professor from Raghu Engineering College. I completed B.Tech in CS & IT and M.Tech in IT. Currently pursuing Ph.D in Machine Learning.

B. Meena, Assistant Professor from Raghu Engineering College. I completed M.Tech in CST. Currently pursuing Ph.D in Machine Learning.

K. Pavan Kumar, Assistant Professor from Raghu Engineering College. I completed B.Tech in ECE and M.Tech in CS & IT. Currently pursuing Ph.D Security issues in multimedia sensor networks, IoT.