BIG DATA ANALYTICS OF GEO-SOCIAL MEDIA FOR PLANNING AND REAL-TIME DECISIONS

RUDRA NARENDRA [#1], L. SOWJANYA [#2]

[#1] MCA Student, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

[#2] Assistant Professor, Master of Computer Applications,

D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.

Abstract

Geosocial Network data can be served as an asset for the authorities to make real-time decisions and future planning by analyzing geosocial media posts. However, there are millions of Geosocial Network users who are producing overwhelming of data, called “Big Data” that is challenging to be analyzed and make real-time decisions. Therefore, in this proposed work, we proposed an efficient system for exploring Geosocial Networks while harvesting data as well as user's location information. A system architecture is proposed that processes an abundant amount of various social networks' data to monitor Earth events, incidents, medical diseases, user trends, and views to make future real-time decisions and facilitate future planning. The proposed system consists of five layers, i.e., data collection, data processing, application, communication, and data storage.

1. INTRODUCTION

To react progressively, associations must channel, connect and process vast volumes of quick moving information in numerous arrangements. Occasion based choice administration frameworks utilize leaders, not simply basic leadership administrations, to react to rapid information continuously. They combine pattern recognition with business decisions and recognize that business decisions need to be shared between event-oriented and process-oriented solutions. Social media are driving their role dramatically day by day as they change from long range informal communication (i.e. social networks) to geo social systems. This has prompted expanded utilization of geo social systems, offering clients the capacity to voice feelings, report occasions and trade perspectives, outrage or love while drawing in with others unbelievable in the pre-web period was.

The data shared over all media is geosocial on the grounds that:

1) The posts have rich substance that speaks to geographic data with specific areas that are either entered expressly (at registration) or included verifiably (by geographic directions, for example, scope or longitude), and

2) Views shared via web-based networking media uncover social information and reinforce connections and correspondence.

With the advancements in the mobile networks and the process of inventing new mobile devices made the people attracted to the location based services very rapidly. The geosocial networking sites have the location based service as distinguish feature. Foursquare which is one of the popular location based networking site enables its registered users to explore the places nearby. Innovative advances have empowered the utilization of GPS frameworks in cell phones, making area information overwhelming. Where individuals post, remark or transfer to web based life is recorded. By amassing such sorts of area information from all system clients, interpersonal organizations hence make geo social information stores. Another technique for creating geo-social information is publicly supporting, while self-created applications are accommodated an assortment of purposes or causes. It recovers geo-social information from volunteers or paid clients who give information or data to this reason. This sort of online information assembled through publicly supporting is named "Volunteer Geographic Information" (VGI) [2].

Today, numerous stages and programming have been produced to use publicly supporting for geo social information collecting purposes to expand business, advance causes, or adventure other business purposes. A case of such a product stage called Ushahidi, which enables reports to be created by coordinating a particular catchphrase in a geointerpersonal organization that compares to particular areas. The report would then be able to be utilized for instruction and help with crises or catastrophes [3].

In this proposed work, we proposed an efficient system for exploring Geosocial Networks while harvesting data as well as user's location information. A system architecture is proposed that processes an abundant amount of various social networks' data to monitor Earth events, incidents, medical diseases, user trends, and views to make future real-time decisions and facilitate future planning.Here we try to extract the information automatically by deploying five layers for data pre-processing.

2. LITERATURE SURVEY

INRODUCTION

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, ten next steps are to determine which operating system and language used for developing the tool. Once the programmers start building the tool, the programmers need lot of external support. This support

obtained from senior programmers, from book or from websites. Before building the system the above consideration r taken into for developing the proposed system.

RELATED WORK

At present, analysts are increasingly intrigued by geo social systems since they think they are new information resources as proposed in the paper [6, 7]. Crooks et al. [8] utilized Twitter information in their work to delineate in the United States. Also, Chow in [9] and Papadimitriou in [10] proposed designs that utilized geo-label suggestions while shaping new informal community. Geo Life 2.0 [11] performed comparability identification, circle investigation.

Table 1: comparative study of literature survey

Done by	Alexis Papadimitriou	Brian Eriksson	Laura Ferrari
Knowledge base	Data collection and recommending users	Location analysis	Pattern analysis
Approach	Higher Order Singular Value Decomposition (HOSVD) technique.	Learning based IP geolocation approach	Topic model based approach
Objectives	To get recommendations on friends ,locations and activities	To pinpoint the geographic location of the IP accurately	To automatically extract urban patterns and recurrent behaviour from location based network
Advantages	New users, locations, or activities can be easily inserted due to the use of tensor algorithm	Need to study machine learning problem is reduced Geolocation estimates were very accurate	Accurately analyses urban patterns
Algorithm/protocol	Tensor algorithm	Naives Bayes estimation algorithm	Latent Dirirchlet Allocation

Table 2: comparative study of literature survey

Done by	Bobadilla .J	Yaronkanza	Yanhua Li* , Moritz Steiner$
Knowledge base	Recommender systems	Geospatial data analysis	Popularity detection
Approach	Content based filtering and model based	intuitive approach	Categorising venues and then predicting popularity
Objectives	To explain the evolution of recommender systems and provide an original classification.	To analyse geo-tagged posts to discover places that were jointly visited by many people	To detect the popular categrie of venues and the predicting popular venue individually
Advantages	Describes the Recommender system trends to implicitly collect data	Improves and facilitate the work of urban planners and of policy makers,	Can know the most popular places in the city and it's descendings
Algorithm/protocol	Bayesians ,genetic algorithms	Clustering ,algorithms, agglomerative clustering	-
Future	Improve the quality of recommender systems predictions and recommendations.	To compare with other parameters	To improve the prediction results to be more accurate
People	Minimal requirement	Required	Minimal

There are couple of different frameworks that perform examination of the tweets, their unique circumstance, e.g. B. [12] - a few employments. Some work has likewise been done to identify hotspots and hyper neighborhood occasions utilizing tweeter information [3, 14] in a city. Social networking organization information could be useful to numerous territories if very much examined. By separating the social direct of a system [15] in an explicit region through filtering and profile planning, one can endorse a store, lodgings, unobtrusive markets, dealing with a record structures, advancements, etc. subject to their tendencies and imperatives [16]. So also, in view of traveller and vehicle developments, open experts can complete better city maps [17]-[19] and prescribe street transport to appropriate street clients dependent on current conditions [20], [21]

3. EXISTING SYSTEM

In the existing system classification of all tweets or messages from geo social network is almost done manually hence we are having a lot of limitations like

LIMITATION OF EXISTING SYSTEM

1. **Time Consuming**: - As this system needs lots of manpower to cross check each and every post which is posted by several users in geo social network like twitter, facebook, linkedin and so on, lots of time is consumed. So time constraints maintenance is very difficult.

2. **Lack of Proper Information**: - All the existing networks try to extract the post based on manual approach, there is lack of proper information to get the updated information.

4. PROPOSED SYSTEM

In this proposed work, we proposed an efficient system for exploring Geosocial Networks while harvesting data as well as user's location information. A system architecture is proposed that processes an abundant amount of various social networks' data to monitor Earth events, incidents, medical diseases, user trends, and views to make future real-time decisions and facilitate future planning. Here we try to extract the information automatically by deploying five layers for data pre-processing.

ADVANTAGES OF THE PROPOSED SYSTEM

The following are the advantages of the proposed system. They are as follows:

1) It is very efficient

2) It is very Accurate for Data Extraction.

3) All the post can be extracted based on keywords which are used and then process the post based on those keywords

This is accurate in classifying a lot of data collected from data sources.

5. SOFTWARE PROJECT MODULES

Implementation is a stage where theoretical design is converted into programmatically manner. The implementation will be divided into number of modules like 4 modules

1. Register Module

2. User Login Module

3. Load Dataset Module

4. Pre-processing Module

Now let us discuss about each and every module in detail as follows:

5.1 Register Module

In this module, the user need to register with all his basic details in order to get login into the account.Once the user is registered then he can login into the account for processing the information.

5.2 Login Module

In this module, he logs in by using his/her user name and password. After Login the user can able to process the pre-processing steps of given dataset.

5.3 Load Dataset Module

Here we try to load the tweets dataset collected from twitter.Which contains almost more than 10000 tweets posted by several users on several topics.Here the data set will be having some columns such as tweet name,posted by,date and time,comments by, and lot more.Based on the important keywords the real time event is extracted.

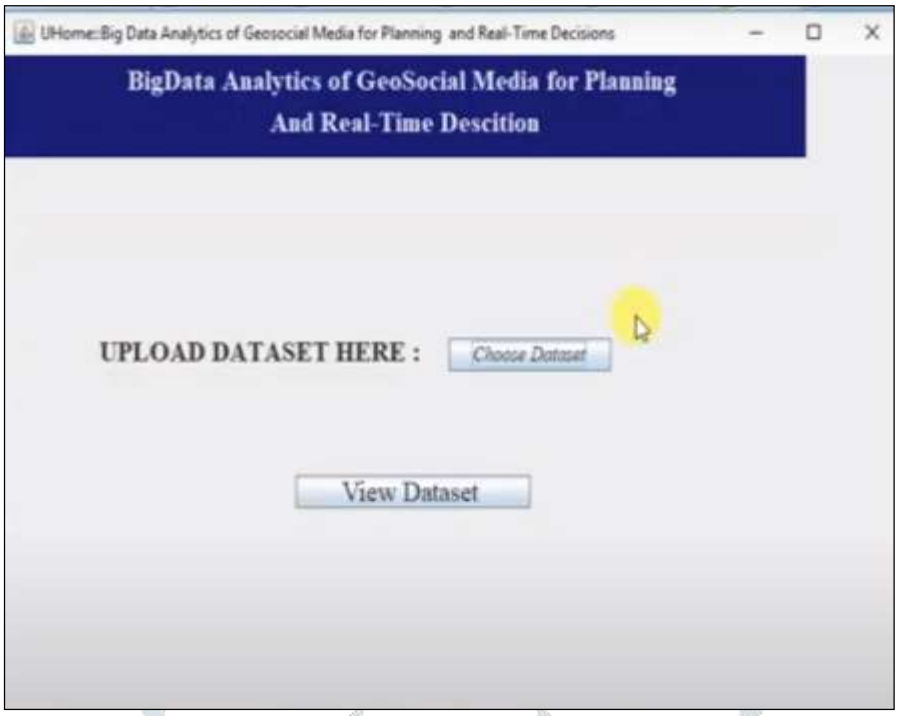
5.4 Pre-processing Module

Here in this module,we try to apply pre-processing technique once the tweet dataset is loaded.Here we try to apply several pre-processing techniques in order to extract the given tweet and find out the real time event from geo social tweets dataset.

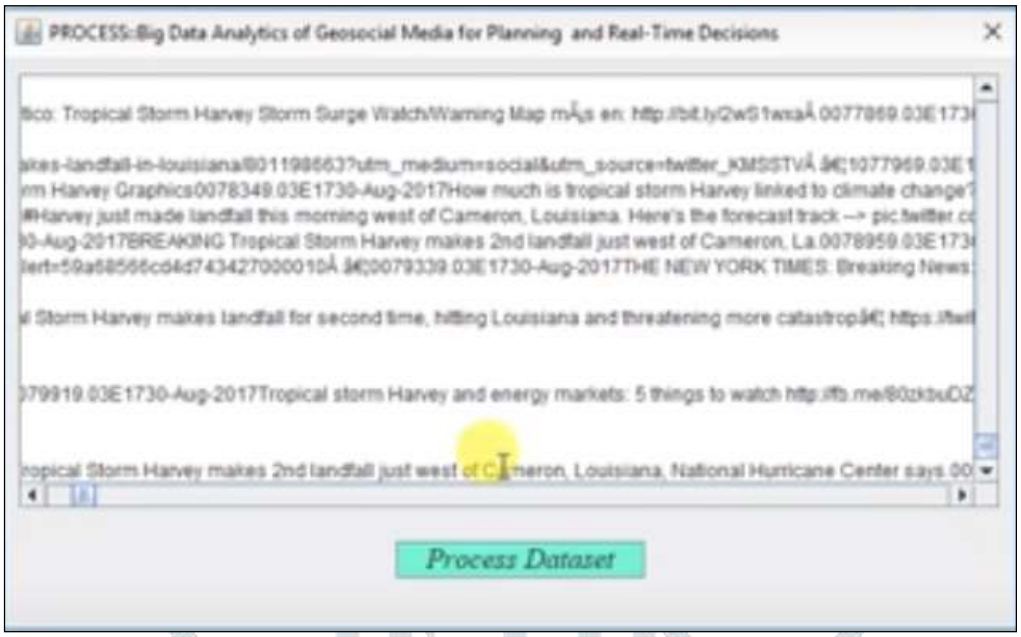
6. RESULTS (OUTPUT SCREENS)

USER LOGIN

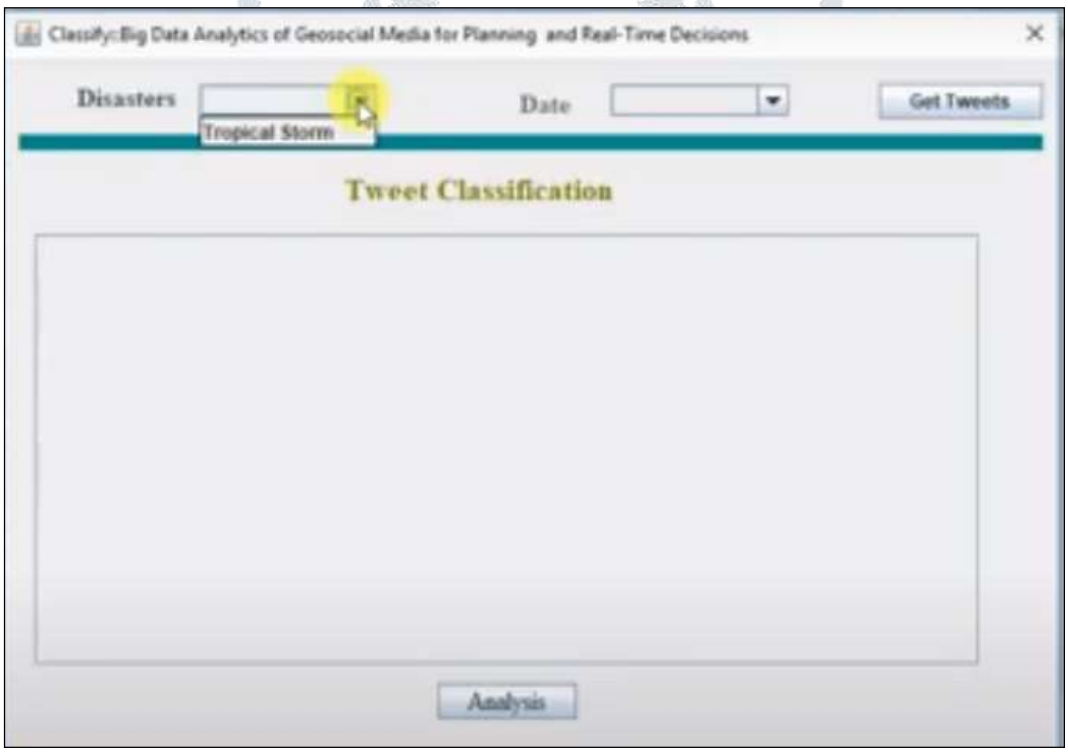
USER MAIN WINDOW



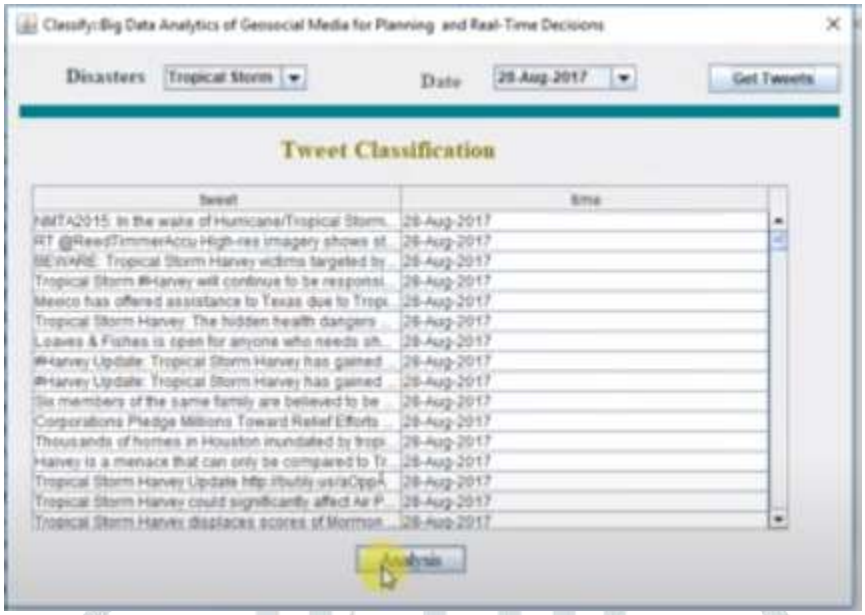
User Upload a Dataset

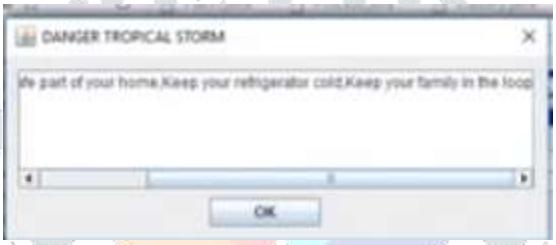


User Process the Dataset

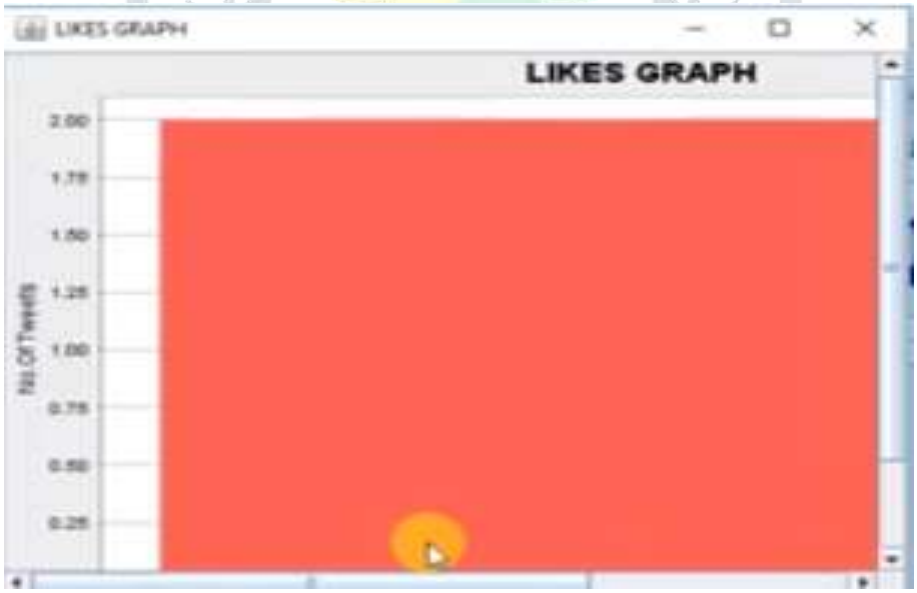


User Tweet Classification





User Can View Lines Graph for Future Prediction



7. CONCLUSION

This software presents the part of an organization work as per the requirements, specifications and conditions mentioned in the user manual. This application s/w has been developed and completed

successfully and also tests successfully by taking “Test Cases”. It is user friendly and has all the needed menu options, which can be utilized by the user to perform the desired operations .Moreover help messages are provided which will help any authorized user from using the system without trouble.

8. REFERENCES

1. HERN, “Online Volunteer Map Philippine after Typhoon Haiyan”,The Guardian, November 15,2013.

2. M. Haklay: “How extraordinary is resolved geographic information? Relative examination of Open Street Map and Ordnance Survey datasets,” Condition and Planning .B, orchestrating and structure, vol. 37, no. 4, p. 682. 2010.

3. M. Zook et.al. “Volunteer Information and Crowdsourcing Disaster Relief: A Case Study on the Haitian Earthquake.” “World Medical and Health Policy,” Volume 2, No.2 ,Pp.7-33,2010

4. Hoot suite, “hootsuite.com”.

5. 140kit, “https://github.com/WebEcologyProject/140kit.”

6. Stefanidis, et.al “Choosing Environmental Information From Social Media Feeds,” Geo Journal, pp. 1-20,2012

7. Z. Cheng, et.al, “Exploring Millions of Suspicious Imprints in SiteSharing Services.”, ICWSM, vol. 2011, pp. 81-88, 2011.

8. Crooks, A. Croitoru et.al. “# Earthquake: Twitter as Distributed Sensor System”, Transactions in GIS, 17 (1), p.124-147, 2012.

9. Chow, J. Bao et.al “Towards Site-Based Social Network Services,” second ACM SIGSPATIAL International Workshop on Location Based Social Networks,pp.31-38,ACM,2010

10. Papadimitriou, P.Symeonidis et.al “Geo-Social Recommendations”, ACM Recommender Systems 2011.Workshop on Personalization in Mobile Applications,2011.

11. Y. Zheng, Y. Chen et.al. “GeoLife 2.0: A Location-Based Social Network Service,” IEEE Tenth International Conference on Mobile Data Management: Systems, Services, and Middleware, 2009 (MDM '09), Pp. 357-358, 2009.

12. O'Connor, M. Krieger et.al “Tweet motiv: Explorative Search and Thematic Summary for Twitter,” in ICWSM, 2010.

13. L. Ferrari, A. Rosi et.al. “Removing urban precedents from region based relational associations”, In Proc. of the third ACMLBSN, 2011.

14. Xia, R. Schwartz et.al. A. Langdon, J. Ting and M. Naaman "City Beat: Real-time Social Media Visualization of Hyper locality Data" In Proc. the WWW Conference 2014.

15. Paul, Anand et al. "Smart Buddy: Defining Human Behavior Using Big Data Analytics on the Social Internet of Things." IEEE Wireless Communications 23.5 (2016): 68-74.

16. J. Bobadilla, F. Ortega et.al. "Recommender systems Survey", Knowledge-Based Systems, 46: 109-132, 2013.