International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 03, March 2016
www.ijates.com
ijates
ISSN 2348 - 7550

REAL TIME TWEET ANALYSIS FOR EVENT DETECTION & REPORTING SYSTEM FOR EARTHQUAKE

Ghansham V. Shendge[1], Mangesh R.Pawar[2], Nikhil D.Patil[3], Pratik R.Pawar[4], Guide: Prof. Devdatta B.Bagul[5]

[1](Computer,BVCOERI,Nashik), [2](Computer, BVCOERI, Nashik), [3](Computer, BVCOERI, Nashik) [4](Computer, BVCOERI, Nashik), [5](Computer Dept, BVCOERI, Nashik)

ABSTRACT

Communication is a key factor in Today's human life, due to time constraints physical interaction between people is not possible. This gap is filled by the technology through 'social networking' sites it's very easy to get access to interact other based on their interests. Many applications are getting releasing with new features day-by-day from vendors, to provide efficient usability and user friendliness.Visualization is a new trend setter of information representation, the back bone of visualization is data. This project proposed a new system that delivers large database of Social Networking Site (SNS) called 'Twitter'. Many Third party application are building based on SNS like Twitter, they need to have processed data from their operational purpose. The main stream of the applications is visualization applications.This project gives more beneficial solution by providing in-depth detailed information of data. In this context this implementation serves processed information of tweets accessed from Twitter Server. Here processing the tweet involves extraction of metadata of tweet, geocoding the physical address in a tweet, analyzing the sentiment of content in the tweet text and extracting the significant and key phrases from text. This application is an integrated system used to get connect and access tweets from Twitter to get processed text analysis components. After all the Information Extracted and NER (Named Entity Recognition) text analysis from tweet, are stored into a persistence database. This document discussed in review the contemporary and early works and studies related to Text analysis for and efficient procedures in extracting vital aspects of information. Here Object Oriented programming and Design patterns are used in implementation of this system, with proper testing and validation are performed at three levels, both normal and performance test results are evaluated to achieve a sophisticated system.

Keywords - Twitter, Event detection, Social sensing element, Location estimation, Earthquake.

I. INTRODUCTION

Twitter is classified as a microblogging service. Microblogging may be a variety of blogging that allows users to send transient text updates or micromedia like images or audio clips. Microblogging services aside from Twitter embody Tumbler, Friend Feed, Jaiku, identi.ca, and others [14].3 Our study, which is S.Anand& K. Narayana International Journal of rising Engineering analysis and Technology ninety-seven based on the period nature of

1 social networking service, is applicable to alternative small blogging services, however we tend to specifically examine Twitter during this study thanks to its quality and knowledge volume [15]. An important characteristic that's common among small blogging services is their period nature. Though web log users usually update their blogs once each many days, Twitter users write tweets many times during a single day. Users will savvy alternative users do and infrequently what they're brooding about currently, users repeatedly come back to the location and check to check what people do. Many necessary instances exemplify their period nature: within the case of a very sturdy earthquake in Haiti, several footages were transmitted through Twitter [16]. Folks were thereby able to apprehend the circumstances of injury in Haiti straight off. In another instance, once associate aeroplane crash- landed on the Hudson River in the big apple, the primary reports were revealed through Twitter and tumbler [17].

This column fine signifies the inspiration of our study. The analysis question of our study is, "can we incline to determine such occasion prevalence in period of time by watching tweets?" This paper presents associate investigation of the period of time nature of Twitter that's designed to determine whether or not we are able to extract valid data from it. We tend to propose an occurrence notification system that monitors tweets and delivers notification promptly victimization information from the investigation. during this analysis, we tend to take 3 steps: 1st, we tend to crawl varied tweets associated with target events First, to get tweets on the target event exactly, we have a tendency to apply linguistics analysis of a tweet. for instance, users may create tweets like "Earthquake!" or "Now it's trembling," that tremor or trembling may be keywords, however users may also create tweets like "I am attending AN Tremor Conference," or "Somebody is shaking hands with my boss." we have a tendency to prepare the coaching knowledge and devise a Tweets Classifier employing logistic regression (KLR) supported options like keywords in a very tweet, the quantity of words, and also the context of target-event words. When doing thus, we have a tendency to acquire a Variation space time model of an instance. We have a tendency to then create a vital assumption: every Twitter user is considered a detector and every tweet as sensory data. These virtual sensors, that we have a tendency to designate as social sensors, square measure of a large selection and have numerous characteristics: some sensors square measure terribly active; others don't seem to be. A detector may well be inoperable or defective typically, as once a user is sleeping, or busy doing one thing else. Consequently, social sensors square measure terribly creaking compared to standard physical sensors. relating to every Twitter user as a detector, the event-detection downside will be reduced to at least one of object detection and placement estimation ubiquitous/ pervasive computing surroundings during which we've various location sensors: a user features a mobile device or a full of life badge in surroundings wherever sensors square measure placed.

II. LITERATURE SURVEY

Twitter is a noteworthy example of the foremost recent kind of social media. Varied researchers have examined Twitter. Relating to similar analysis to it conferred during this paper, some researchers have tried topic detection victimization Twitter. Cataldi et al. projected a unique technique to discover rising topics employing a keyword-based topic graph. They succeeded in detective work news keywords that area unit fashionable in Twitter. As an example, (a volcano in Iceland) and Samaranch (the previous President of IOC, World Health Organization died

in Apr 2010). Marc et al. divided more and more fashionable keywords on Twitter into patterns of assorted type's victimization Kyrgyzstani monetary unit, thus signifying that Tweet users add to the discussion of those trends. Other than the studies introduced in Section one and these studies, many others are done. We have a tendency to classify studies coping with Twitter or knowledge on Twitter into 3 teams. First, some researchers specifically examine the network structure of Twitter and investigate Twitter network options of assorted types. Java et al. analyzed Twitter as early as 2007.

We gift a quick summary of Twitter in Japan: The Japanese version of Twitter was launched on Gregorian calendar month 2008. In February 2008, Japan was the No. a pair of country with relation to Twitter traffic.5 At the time of this writing, Japan has the second largest range of tweets (18 % of all tweets are announce from Japan) within the world. Therefore, we decide earthquakes in Japan as a target event as a result of the high density of Twitter users and earthquakes in Japan.

III. EXISTING SYSTEM

The Existing system, called Toretter is presently working in Japan for Earthquake detection using Twitter has been operated since August 8; 2010.Users can see the detection of past earthquakes. Also they can register to receive notices of future earthquake detections. It alerts users for imminent earthquake. It is hoped that a user receives alert before the earthquake actually affects the area. We assess various conditions under which alarms might be sent to choose better framework for our suggested system.

IV. PROPOSED SYSTEM

The reference systems used reduced corpus datasets that do not scale to larger amounts of data. The performance, effectiveness and durability of those systems was not being designed to handle big amounts of data but today's online social network service volume of data (creates massive unstructured text data streams) make them obsolete systems. Performing real-time event detection using Twitter requires dealing and mining massive un- structured text data stream that has messages continuously approaching at sky-high data rates. Given this, the approach to deal with this specific problem involves providing solutions that are able to mine continuously, high-volume of open-ended data streams as they arrive. Considering that those sources of data are coming from social network users it is expected that information collected using metrics of networks analysis (nodes, connections and relations, distributions, clusters and communities) could improve the quality of the solution of the algorithm. Apart from, the data in online social network services is also dynamic; messages and arriving at very high data rate. Computation of such vast amount of data needs necessarily technology that has a highly scalable storage platform and performs distributed concurrent parallel execution of database. Time and Cost Effectiveness is an issue. Online social network text streams seem to be the ideal source to perform real-time event detection as they are very much Cost Effective.

4.1 Algorithm

1) Developing and Preparing Datasets

We are trying to come up with a function that can predict for future inputs based on the practice it has acquired through the past inputs and their outputs (training set).

2) Logistic Regression is - coming up with a possibility function that can give's a chance, for an input to belong to any one of the various classes'(Classification)

3) Consider binary classification problem as each tweet is either positive (y = 1) or negative (y = 0). These are the 2 parameters here. Our goal is to come up with a possible function that takes in an input X (Number of tweets) and return 'what is the possibility of this tweet to be positive'.

4) This probability function is the Sigmoid Function and which is:

$$\frac{1}{(1+e)^{(-z)}}$$

Since, probability of any occurrence is [0, 1] (between 0 and 1, including both), this task absolutely fit to be used as a probability function for logistic regression.

z = transpose (theta) * X X =Number of Tweet Theta=?

X = 0.9 and it gave probability for it to be positive = 0.3, which means it has more possibility of starting gently. But clearly from our training set this is definitely wrong as for X = 0.9, Y = 1 i.e. malignant.

5) In general the error function in logistic regression is given by:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} cost(h_\theta(x^{(i)}), y^{(i)})$$

Here m = no of elements in training set,

Y is commonly 1 or 0 and h(x) is solely the 'Sigmoid function' we spoke above. Since sigmoid is a function of theta (specified in z above), therefore J is a set of theta.

6) Now we minimize J over theta and search out those values of theta for which our error function in minimized.

7) Once we have theta, our probabilistic result (sigmoid) is ready and we can apply it to any size of data and it will give us its.

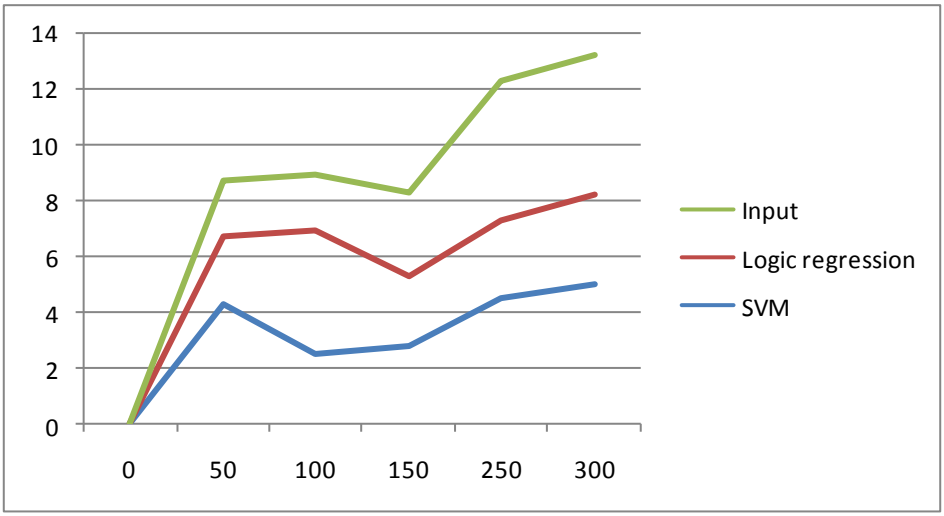
V. RESULT ANALYSIS

ALGORITHM DATA SET RESULT ANALYSIS:

Parameter	Data Set (X axis)	Input Vector Data Set	Support Vector Mechanism (old Algorithm)	Logic Regression Classifier (New Algorithm)
	0	0	0	0
	50	8.2	4.1	6.7
Algorithm Performance	100	8.4	2.2	6.2
	150	8.1	2.4	4.2
	250	12.1	4.2	7.2
	300	13.6	5	8.1

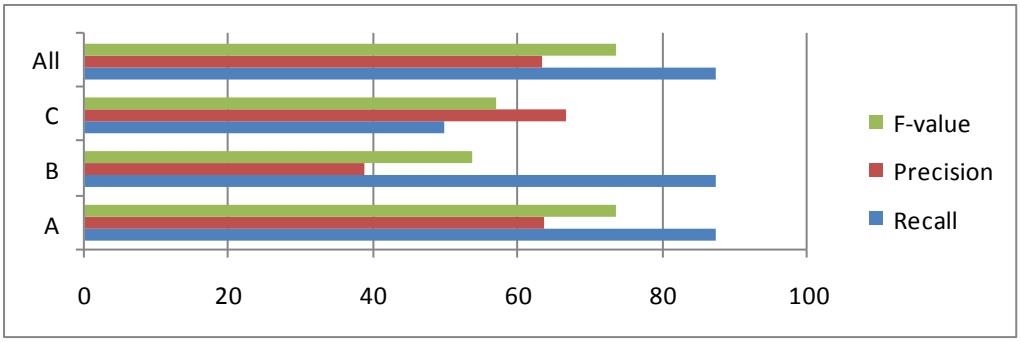
Result will be generated in graphical chart; here this chart is generated from using Google chart. This chart shows the positive and negative polarity. This is generated by using GATE processor's score of comment associated with particular product.

By using this chart user can analyze product efficiently. GATE processor performs major task in generating scores of an comment. These scores such as negative and positive is generated from comment.



Performance analysis

Recall	Precision	F-Value
87.50	**63.64**	**73.69**
87.50	**38.89**	**53.85**
50.00	**66.67**	**57.14**
87.50	**63.64**	**73.69**



International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 03, March 2016
www.ijates.com
ijates
ISSN 2348 - 7550

OVERALL SYSTEM WORK FLOW:-

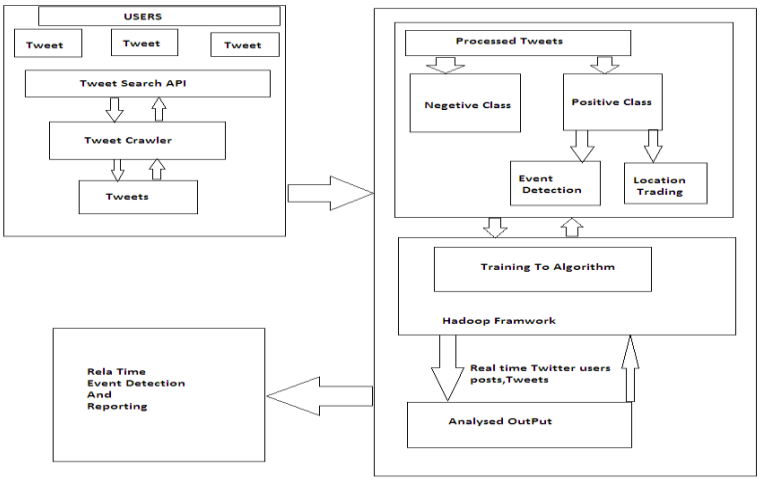
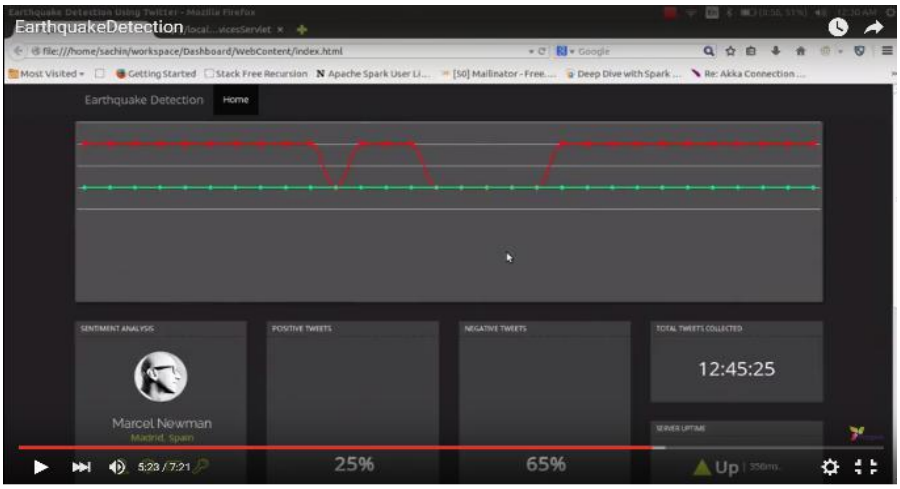
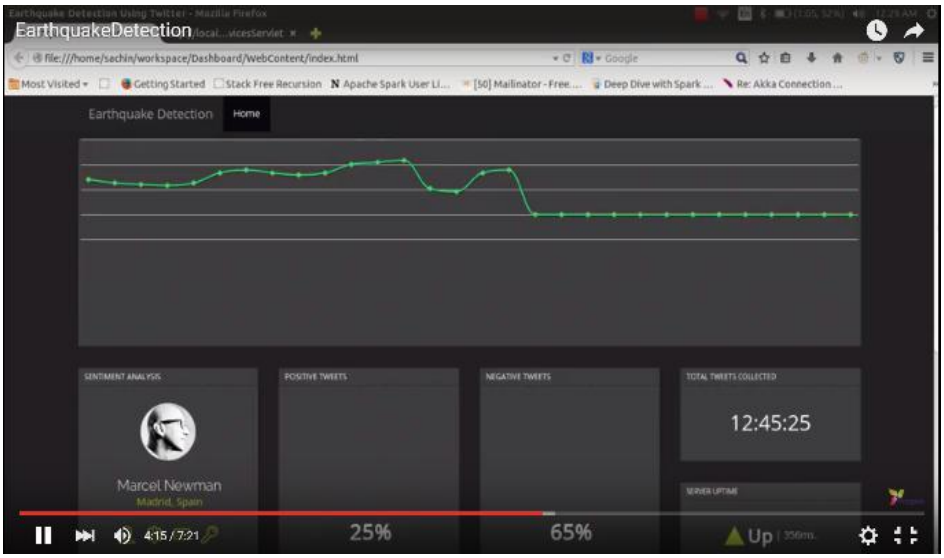
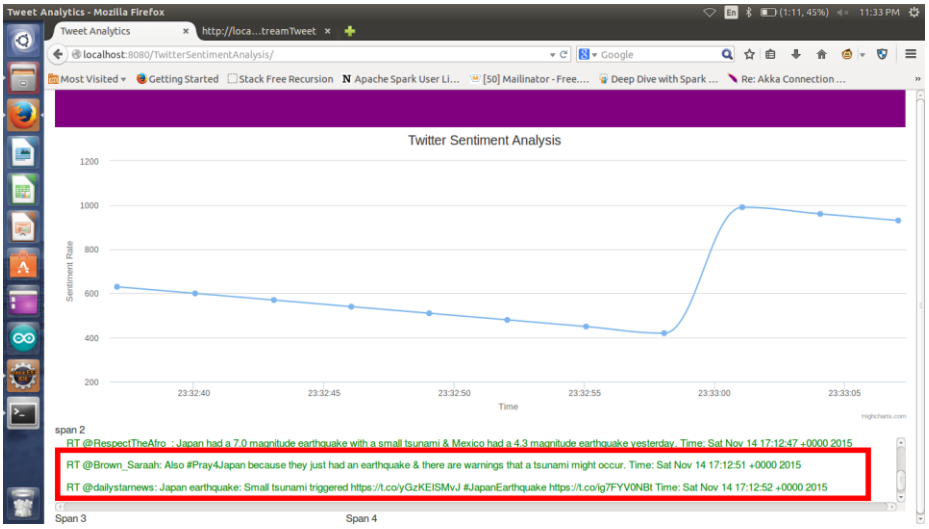


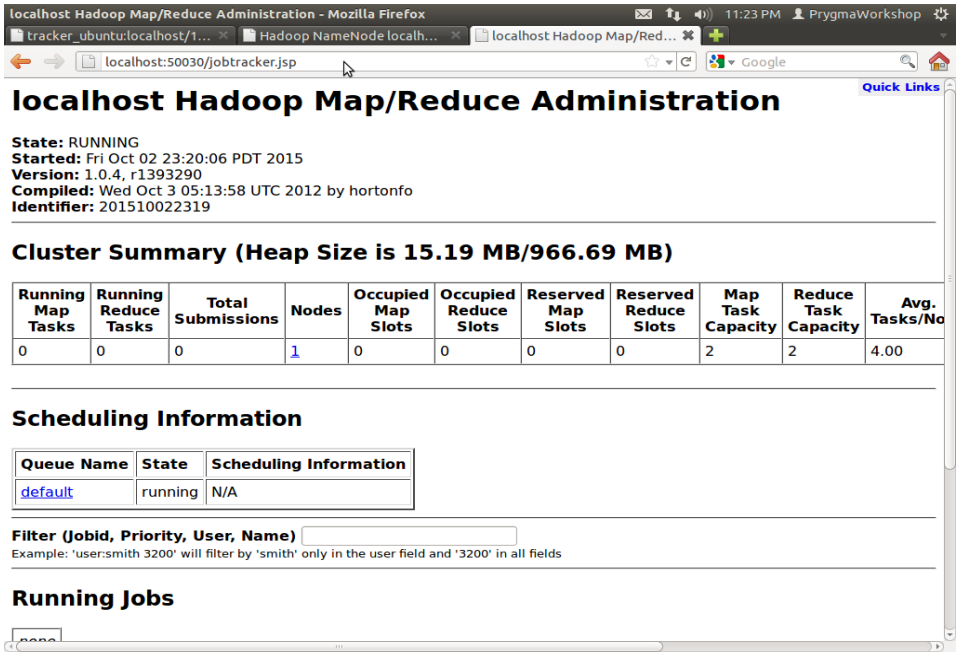
Fig. System work flow





International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 03, March 2016
www.ijates.com
ijates
ISSN 2348 - 7550





VI. CONCLUSION

We have a tendency to investigate the period nature of Twitter, devoting specific attention to event detection. Linguistics analyses were applied to tweets to category verifies them into a positive and a negative class. We have a tendency to regard every Twitter user as a device, and set the matter as detection of a happening supported sensory observations. Location estimation strategies like particle filtering area unit are used to estimate the locations of events. As associate degree application, we have a tendency to developed associate degree earthquake coverage system, which could be a novel approach to advice folks promptly of associate degree earthquake event..

ACKNOWLEDGMENT

This work was supported by Pune University and BVCOERI, Nashik. We are very much thankful that he gave opportunity to complete this work in time to us. We would also like to thanks our Prof. C. K.Patil Principal BVCOE&RI. Prof. H. D.Sonawne H.O.D Computer Department, for providing their valuable support and time throughout engineering.

REFERENCES

[1] G. V. Shendge, M. R. Pawar, P. R. Pawar, N. D.Patil, Prof. D. B. Bagul "Real time Tweet analysis for event detection & reporting system for Earthquake" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2395-0056 Volume: 2 Issue: 2395-0072.

[2] A. M. Sarah, C. Abdur, H. Gregor, L. Ben, and M. Roger, "Twitter and the Micro-Messaging Revolution," technical report, O'Reilly Radar, 2008.

[3] Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proc. Ninth WebKDD and First SNA-KDDWorkshop Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07), pp. 56-65, 2007.

[4] B. Huberman, D. Romero, and F. Wu, "Social Networks that Matter: Twitter Under the Microscope," ArXiv E-Prints, http://arxiv.org/abs/0812.1045, Dec. 2008.

[5] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, A Social Network or A News Media?"Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 591-600, 2010.

[6] Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment," Proc. Fourth Int'l AAAI Conf. Weblogs and Social Media (ICWSM), 2010.

[7] P. Galagan, "Twitter as a Learning Tool. Really," ASTD Learning Circuits, p. 13, 2009.

[8] K. Borau, C. Ullrich, J. Feng, and R. Shen, "Microblogging for Language Learning: Using Twitter to Train Communicative and Cultural Competence," Proc. Eighth Int'l Conf. Advances in Web Based Learning (ICWL '09), pp. 78-87, 2009.

[9] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," Computer, vol. 34, no. 8, pp. 57-66, 2001.

[10] M. Weiser, "The Computer for the Twenty-First Century, "Scientific Am., vol. 265, no. 3, pp. 94-104, 1991.

[11] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development," Proc. Int'l Conf. World Wide Web (WWW '10), pp. 9526, Vol 25 April 2013.

[12] Krūms, Jānis (January 15, 2009). "There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy".