# A NOVEL APPROACH TO TWEET ALERT FOR IDENTIFYING REAL TIME TRAFFIC

**[1]P. Sunil Kumar,    [2]Polisetty A Satish Kumar ,    [3]M Anjineyulu,    [4]Boya Talari Panduranga**

[1,2,3]Assistant Professor, [4]Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

## ABSTRACT

At present, one of the major issues for an individual to meet to their requirements is the disordered traffic. In order to resolve this issue, this paper focusing on designing an application to assign the true class label for each and every individual tweet related to the traffic words. If any message, which contains traffic related information, it is being sent as an alert to the end users who are following the current user, and else the same tweet will be just posted on the user wall. In digital era, the social networks have become a fascinating domain for every human for sharing and communicating their recent updates among each other. In order to implement this application, it chooses a compatible social media that is twitter.

**Keywords:** Tweet Classification, Social Networks, Text Mining Technique, Twitter Stream Analysis.

## INTRODUCTION

A popular small blogging service, has received much attention recently is the Twitter.This is one of the OSN service which attained a variety of individual's interest towards it in terms of updated status information can be shared among the friends, family and coworkers [1].In twitter each and every message is termed as a status update message or simply SUM, which is just a message to wish friends or colleagues. Till now there were a lot of researchers who have printed their studies of Twitter to date, especially throughout the past year. Most studies can be classified into one in every of 3 groups: In the first group, we came to know that maximum researchers try to analyze the complete network structure of Twitter and they want to calculate the work load that is present for the twitter application [2], [3], [4]. In the second group, many researcher staff tries to examine orfind out the important characteristics of Twitter application as one of the social medium [5], [6]. Finally, the last group users try to create new apps to compete the twitter [7], [8].

Twitter is categorized as a small blogging service. Micro blogging is a kind of blogging that allows users to send temporary text updates or micro media like images or audio clips. Micro blogging services alternative than Twitter embody Tumblr, Plurk, Jaiku, identi.ca, and others.
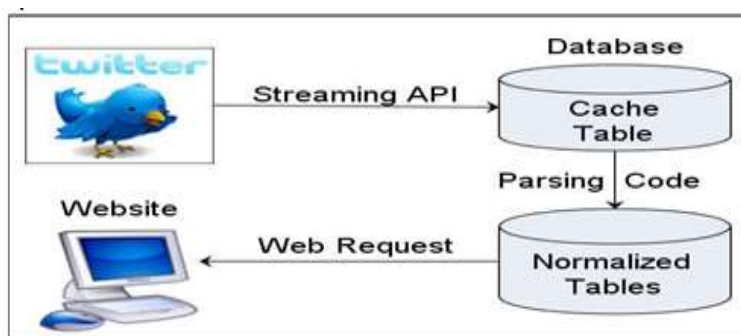


Figure.1: Represents the Sample Architecture of A Twitter

The architecture of twitter is explained in detail in the above figure 1, where each and every twitter has a facility to stream the information. For this twitter will connect with stream API, once the twitter sends any tweet it will be saved into the database with the help of cache table. Now the information which is available in the cache table will be parsed, the term parsing indicates that token is generated for the each and every individual tweet. In this parsing approach it will try to normalize the tweets and eliminate if thereare any

duplicate tweets that are available in the cache table. Once the tweets are normalized then the information is passed into the website in the form of a web request.

Generally, in the twitter there are a separate terminology that was used in order to represent the tweets, one among them is when a user message is shared in the social networks,itiscalledasStatus UpdateMessage (SUM),andit maycontain,apart fromthe text, it also contain a other information like timestamp, geographic coordinates (i.e. latitude and longitude), username who posted that tweet and also the has tags that was applied for that appropriate tweet. As we all know that a lot of SUMs which refer to a certain topic or related to a limited geographic area may provide some information about an event or topic [9]- [15].

## RELATEDWORK

In this section we mainly discuss about the knowledge regarding the working of twitter and its various services. Now let us discuss about those in detail as follows:

### Motivation

Now a day's twitter is the main source for spreading the ideas and information throughout the Web from one location to other. The message which was passed through twitter is known as tweet, where it will mainly discuss about various trends, innovative ideas, upcoming or ongoing events, and a lot more. This gave rise to an increasing interest in analyzing the various tweets by the data mining community. In current days a lot of OSN users, try to use the twitter as a good resource provider for detecting events in real- time. In this current paper author mainly presented a 4 challenges of tweets like:

1.        Health Epidemics Identification,

2.        Natural Events Detection,

3.        Trending Topics Detection, & Sentiment Analysis.

All the above 4 challenges are mainly based on Data Mining principles like clustering and classification. We review all these approaches by providing a detailed description for each and every challenge individually. As we know that twitter is one type of micro blog, their rates of activity reached for a high level without the need of precedent factor. By conducting a survey on the micro blog service, we got an estimation report that over a hundreds of millions of users are registered in this micro blogs as tweet users for communicating and sharing their mutual interest one among other with their last thoughts, different moods or activities in various words. This can be clearly explained in below Fig. 2, where all the four challenges of tweet are shown in tweet event detection model.



Figure 2: Represents the Four Challenges for Tweet Event Detection

## KNOWLEDGE DISCOVERY

Inthissectionwemainlydiscussabouttheknowledgediscoverythatwasusedin proposing and implementing this current paper

1. A Novel Graph Mining technique to show the communication details with thehelp of exchanged

messages.

2. ANovelTextMiningtechniquecalculatedbasedontypeofmessagecontentthatis available during communication.

From the above two contexts, first one is nothing but graph mining based on the analysis of links that are having between each and every tweet messages. Next if u go with an alternate way, there u can find out the text mining approach based on the analysis of the message content what that was messaged by the tweet users. In general, almost all the users try to show their interest on the tweets content using text mining. For describing the event or a tweet, we have four main dimensions' like

a. **Event Type**: what is happening in that tweet?

b. **Time**: when an event is happening?

c. **Location**: where an event is happening?

d. **Entities**:who is involved in an event?

All the above four dimensions are needed for classification or clustering of any tweet. As we told from figure 2therearefoureventswhichtakeplacefora tweet, all the four events mainly deal with the above dimensions for identifying the tweet and its usage.

**A Novel Proposed Traffic Detection System by Using Twitter Stream**

**Application:**

In this section we will find out the Novel proposed traffic detection system by using twitter as the main application including the service like Service Oriented Architecture (SOA) and based on the Status Update Message (SUM) that was used to identify the status of any tweet.

**Preliminaries:**

In this section, we mainly discuss about the proposed traffic detection system based on Twitter streams. The proposed system architecture is a service-oriented and event-driven, and is composed of three main modules, namely:

I.  To identify the Status Update Message (SUM) an do pre-processing.

II. To elaborate the Status Update Message (SUM) and find its importance.

To Classify the Status Update Message (SUM).

From the above three main modules we can observe one thing common in all of the 3 modules like Status Update Message (SUM). These can be shown clearly in figure 3 and this was applied on an example of traffic event identification.
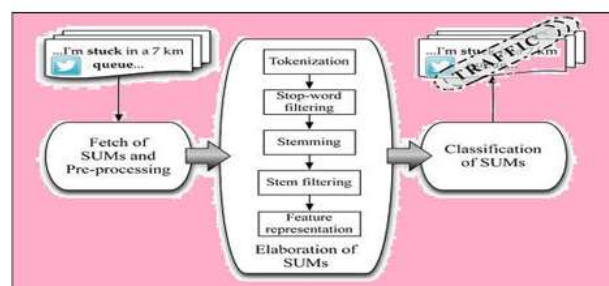


Figure.3: Illustrates the Working Flow Of Proposed Tweet Alert System

**To Identify the Status Update Message (Sum) and Do Pre-Processing:**

In this module we mainly discuss about the status update messages (SUM) and try to identify all the SUM's from all users and try to apply pre-processing for those tweets. Here the term pre-processing in data mining indicates cleaning of messages if there is any abusedcontent,spellingmistakes,extravalues,andspecialcharactersandsoon.Theterm pre-processing will identify all these things and try to ignore un-necessary constants that are available in the messages and try to identify the important characteristic like traffic related keywords. Here each and every tweet contains the information with following fields like: the user id, the timestamp, the geographic coordinates, a re-tweet flag, and the text of the tweet.

**ToElaboratetheStatusUpdateMessage(Sum)andfindits Importance:**

In this module we mainly discuss about the "Status Update Message (SUM) and try tofinditsimportancebyreadingthe SUM".Here weappliedtext miningtechniquein which all repetitive words along with stop words are removed and only the words which contain useful information is only kept available in the SUM. Now we try to arrange all the available words in a set and then identify each and every word with the pre-defined setthat is available in the database. If the available word is matched with the set of words, then such a tweet is known as traffic related tweet and if the same words are not matched with anyof theset,thenthetweets istreatedasnontraffictweet. Here wedivide eachand every tweet into two tokens one is traffic related tweet token and other one is non traffic related tweet token.
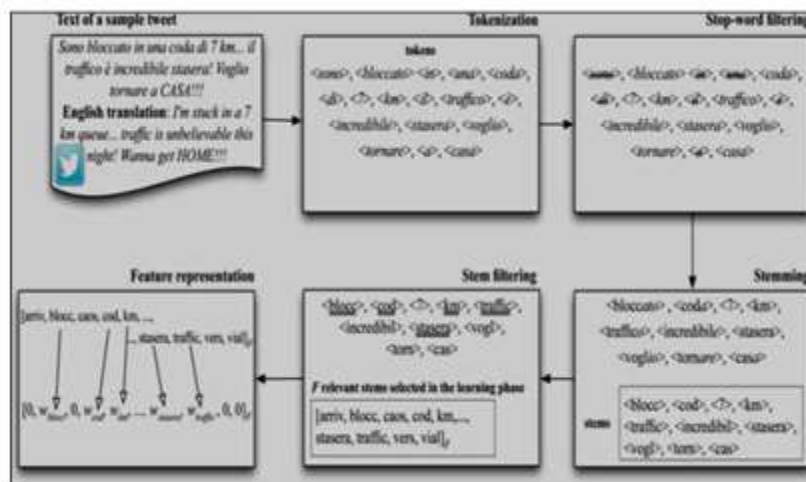


**Fig4:Steps Of The Text Mining Elaboration Applied To A Sample Tweet.**

Fromtheabovefigure4,we can clearly find the clear explanation about the second module with a small example.

**To Classify the Status Update Message (Sum)**

In this module we tryto classifythetweets based on the semantic meaningof eachand everytweet, basicallytweets are classified into twotypes: One isTraffic Tweet and other is Normal Tweet. So basedon the semantic meaningof each and everyuser posted tweet, the tweets are classifiedinto either any one of the twocategories. If the message contains one or more than one traffic related keyword, then such a tweets are classified as Traffic relatedtweetsandifthesame messagecontainsnosinglewordrelatedtotrafficthensuch a tweets are labeled or classified as Normal tweets.

**Procedural Approach of Proposed Model**

Inthissectionwemainlydiscussabouttheproposedmodelthatwascarriedoutinthe current paper. Now let us discuss about that in detail as follows:

LetSistheWholeSystemConsists: S= {I, P, O}
I=Input.

P=Procedure. O= Output.
I= {U,T, TS,URL, Tk}.

1. LetUissetofnumberoftwitterusersinthesys-tem. U = {u1, u2,……un}.
2. TissetofnumberTwittorstatusupdateoftwitteruser. T= {t1, t2, t3…tn}.
3. TSistwitterstreamerwhoanalyzesthe twits.
4. URListheURLoftwitteruser whohaveupdatedstatus.
5. TkisthetokenizationofSUMwhere,SUMistheStatusUpdateMessageof twitter user.

**RESULT ANALYSIS**

As we all are facing much road traffic and in turn a number of accidents while we try to travel from one place to another place. Although there are number of traffic regulatory devices and rules, still there was a heavy traffic jam in current days through which wastage of fuel during waiting in traffic. So in this paper we implement a new tweet system which can monitor the current road traffic and it can send that information to others who are directly linked with that tweet account. This is mainly done with the help of a Twitter stream analysis.

The above figure represents the home page of twitter alert system where we need to signup if we are a new user and login if we already exist as a twitter user.



Fig:4.2 shows the user login page

The below figure shows how friends are being followed and added into the twitter account for updates related to traffic.



**Fig:4.3 Shows the friends added into the account**

Fig:4.4 Show the final tweets showing the alert messages of traffic

The above figure shows the final tweets and the result about final traffic alerts in the city where you can clearly find the areas with traffic problem being notified as an alert.

## CONCLUSION

This paper, described for the first time a novel system for real-time detection of traffic- related events from twitter stream analysis. This proposed system allows us to built a traffic alert system using twitter as a main source for updating the status of traffic to each and everyone who come in that region. Here it can be able to identify if a traffic event is occurred due to real traffic related situation or sometimes it may cause due to some external factors like cricket or football match, procession and manifestation, or not. By conducting various experiments on our proposed model we finally came to a conclusion that our proposed approach is best in providing security for the current society in terms of alerting when there is a traffic near that way and avoiding huge maintenance for the end users who wish to travel in the same direction.

## REFERENCES

1. 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. 2008.4

2. 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.

3. M. Sarah a well known author who proposed a paper on "Twitter and the Micro- Messaging Revolution," 2008.

4. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proc. Ninth WebKDD.

5. 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. for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

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

7. K. Borau, C. Ullrich, J. Feng, and R. Shen, "Microblogging for Language Learning: Using Twitterto TrainCommunicativeandCulturalCompetence,"Proc.EighthInt'l Conf. Advances in Web Based Learning (ICWL '09), pp. 78-87, 2009.

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

9. G. Anastasi et al., "Urban and social sensingfor sustainable mobility in smart cities," in Proc.

IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, Palermo, Italy, 2013, pp. 1–4.

10. A.Rosietal.,"Socialsensorsandpervasiveservices:Approachesandperspectives," in Proc. IEEEInt. Conf.PERCOM Workshops, Seattle,WA, USA, 2011, pp. 525–530.

11. T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real-time event detection andearthquake reportingsystemdevelopment,"IEEE Trans.Knowl. Data Eng.,vol. 25, no. 4, pp. 919–931, Apr. 2013.

12. A.Gonzalez, L. M. Bergasa, and J. J. Yebes, "Text detection and recognition ontrafficpanelsfromstreet-levelimageryusingvisualappearance,"IEEETrans.Intell. Transp. Syst., vol. 15, no. 1, pp. 228–238,Feb. 2014.

13. N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom, and P. Chaovalit, "Social-based traffic information extraction and classification,"in Proc. 11th Int. Conf. ITST, St. Petersburg, Russia, 2011,pp. 107–112.

14. P.M.d'Orey and M.Ferreira, "ITS for sustainable mobility: A survey on applications and impact assessment tools," IEEE Trans. Intell. Transp.Syst., vol. 15, no. 2, pp. 477–493, Apr. 2014.

15. K. Boriboonsomsin, M. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4,pp. 1694–1704, Dec. 2012.