

# Dynamic Road Traffic Detection from Twitter Stream

Prof.V S Karvande<sup>1</sup>, Ms. Sonal N. Gamey<sup>2</sup>

Asst Professor<sup>1</sup>, P.G. Student<sup>2</sup> Department of Computer Science & Engineering, Everest College of Engineering & Technology, Aurangabad, Maharashtra, India  
sonalgamey@gmail.com

**Abstract—** Traffic congestion is a most huge problem all of us are facing nowadays. Peoples are suffering in terms of money and time. In this paper, we present a system to dynamically analysis road traffic and its causes, we know that there are various Popular social networking sites to perform real-time detection of traffic-related information. In this paper we are using Twitter from that twitter, collecting traffic and non traffic related tweets, automatically Classified traffic-related messages posted on Twitter, among the millions of unrelated messages posted by users. on Twitter, From that traffic related messages easily analysis traffic on particular area and reduces the traffic. by detection of real-time traffic events from Twitter stream and traffic & gives that real time information on Television, FM as a news bulletin, because of this traffic is reduces and control the crowd on the road.  
**Keywords:-** Component, formatting, style, styling, insert.

## I INTRODUCTION

Now a day's people using various Social networking sites such as Twitter and Face book, YouTube, google+, whatsapp, hike etc, this sites are the most popular tools for people to share their views opinions, thoughts, feeling on various topics. In this paper the, Twitter is the most popular fastest-growing micro blogging service Twitter was created in March 2006 by Jack Dorsey. Dorsey published the first Twitter message at 9:50 PM Pacific Standard Time (PST) Messages posted on Twitter (tweets) having everything from day to day life stories to the latest local and global news ,ideas and events. In June 2012.1 over 400 million tweets per day with more than 140 million users Twitter enables users to post status updates, or *tweets*, no longer than 140 characters message to *followers* using various communication services (e.g., cell phones-mails, Web interfaces, etc. Proposed system is an intelligent system based on text mining and Natural language processing algorithm, for dynamically detection of traffic events from Twitter stream & it gives that real time information on

Television & FM as the news bulletin[4].we also focus on exact location of user, not the region In this paper, we build an efficient, scalable system to detect events from tweets There are many applications, that uses Twitter for earthquakes, Another category of events detected by Twitter is sports. During., Twitter is also one of the unique online social networking service that allows people to create profiles, communicate, and connect with other people on the service.

The social relationship on tweeter is user can follow any other user without requiring an approval or a reciprocal connection from the followed users. Twitter does not restrict any limits on the number of followers to a user account; In this, one user account can typically follow up to 2000 users [2].The user message shared in social networks is called Status Update Message (SUM), and it may contain, apart from the text, meta-information such as timestamp, geographic coordinates (latitude and longitude), name of the user, links to other resources, hash tags, and mentions. Various SUMs referring to a certain topic or related to a limited geographic area & great deal of valuable information about an event or a topic. SUMs as sensor information, the most challenging problem than event detection from traditional media like blogs, emails, etc., where texts are well formatted [2]. In fact, In the Twitter SUMs are unstructured and irregular texts, that contains informal or abbreviated words, misspellings or grammatical errors According to Pear Analytics,[1] with no useful information for the audience.[3] SUMs contain a huge amount of not useful or meaningless information like as, unstructured and irregular texts, abbreviated words, misspellings or grammatical mistakes, which has to be filter out , For that reason, the proposed system to analyze the information coming from social networks like Twitter, we apply the text mining techniques, It is the field of data mining, machine learning, statistics, and Natural Language Processing (NLP) to extract meaningful information. Natural Language Processing *NLP* refers to AI method of communicating with an intelligent systems using a natural language such as English. But the most challenging problem is storing and processing of large amount of unstructured data, handling high amount of data & also removing noise from abnormal data, and finding expected data or meaning full results In this paper, we proposed an architecture that enables an effective storage and analysis of unstructured data, and perform

the semantic analysis on the data. By using the Natural language processing (NLP) The NLP evaluation is conducted through stemming filtering, Feature representation, System Features on Twitter data, and performing sentiment analysis. [7]. Text mining contains the process, of structuring the input text & recognized patterns for the structured data, and finally evaluation and validation of the output. The text mining is the process particular goal, such as: i) semantic analysis through use of NLP technique, ii) text filtering by means of specific keywords, iii) feature extraction, i.e., conversion of textual features (e.g., words) in numeric features (e.g., weights), that a machine learning algorithm is able to process, and iv) feature selection, i.e., reduction of the number of features in order to take into account only the most relevant ones. for real-time detection of traffic events from Twitter stream and traffic & gives that real time information on Television, FM as a news bulletin, because of this traffic is reduces and control the crowd on the road.

### II LITRATURE SURVEY

This paper describe the design and development of system by using Stanford Core NLP is one of the most used, & identify the attributes that support to its output comes from twitter[5].In the previous system likes Sakaki and others that uses social sensor to sensing real time messages from Twitter & the sensors gives earthquake related events. [8] Another system was Asur and uber man use to predict real-world outcomes. From Twitter, as focuses on the prediction of box-office revenues for out-coming movies.[9] Tumasjan and others focusing in the predictions of Elections. They conclude that the mere number of tweets reflect voter preferences and comes close to election polls, and that the tweets are not only about spreading political opinions, but also to discuss these opinions with other users[10] This system is able to detect, with high probability, most of the earthquakes, small-scale event detection, like Agarwalet al. [7] focus on the detection of fires in a factory from Twitter stream analysis, by using standard NLP techniques and a Naive Bayes (NB) classifier, to detect and analyze small-scale incidents, such as fires. Li et al. [3] propose a system, called TEDAS, to retrieve incident-related tweets.

### III DIAGRAM DESCRIPTION

- 1) The user Tweet on twitter API window collects all the traffic and non traffic related tweets in large scale.
- 2) Here traffic and non traffic related tweet from twitter API are stored in data set.
- 3) Take any tweet from data set and perform tokenization assigns a tokens and removing the any hash tag present in the tweets.

- 4)After the hash tag removing system perform parts of speech (POS) tagging means to assigning the noun, pronoun verb adverb i.e NN, NP, NNP etc.
- 5)we are using Stanford NLP classifier for recognition of named entity from tweet.
- 6)After NLP classifier Processed to distinguished between positive tweet(+) class and negative tweet (-) class”
- 7) From positive class we find out event detection precision recall fmeasure and exact location of user not the region.
- 8) Lastly we improve an actual time tweeter operator’s method to report real time event detection and analysis of traffic reporting.

### IV SYSTEM ARCHITECTURE

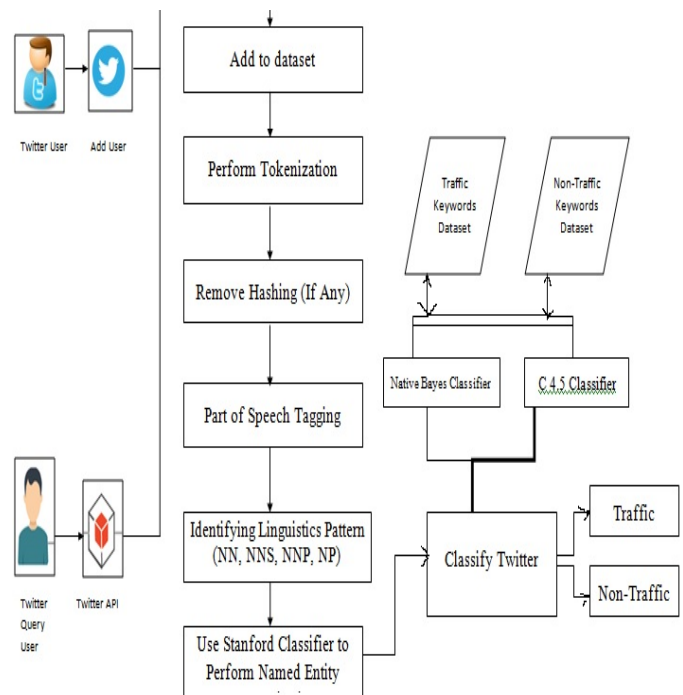


Figure 1 Dynamic Road Traffic Detection From Twitter Stream

### V EXISTING SYSTEM

They have tried to build a meaningful set of traffic events, related to some major Italian cities Twitter’s stream fetched the data from twitter API and uses the WEKA classifier, and the difference between these two times, positive differences indicate a late detection with respect to official news web sites, system exploited available software packages and state-of the-art techniques for text analysis and pattern classification, have shown the superiority of the SVMs, which have achieved accuracy of 95.75%, for the 2-class problem, and of 88.89% for the 3-class problem, in which also considered the traffic due to external event class. The best classification model has been employed for real-time monitoring of several areas of the Italian road network

**VI PROPOSED SYSTEM**

Existing system was developed for Italian road ways not for the Indian roads. We are going to design the system called dynamic road traffic detection from twitter stream based one classifier NLP and C4.5 to process huge amount of tweeter data related with ‘Road traffic’ because of this we improved the accuracy of our system. Actually it is one of much needed project started by seeing people facing traffic problem in India. Due to heavy traffic. Many people lost their lives, and facing accident problems because of this we are proposing such system which will make public announcing on web by analyzing huge In information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also called sensitivity) is the fraction of the relevant instances that are retrieved. Precision and recall are therefore based on understanding and measuring relevance.

**Table 1 Result of traffic**

Precision	Recall	F-measure
99.0566037735 8491	66.8789808917 1973	79.8479087452 4714
96.7032967032 967	67.1755725190 8397	79.2792792792 7928
72.9106628242 0749	67.1087533156 4988	69.8895027624 3095

In simple terms, high accuracy means that an algorithm returns significantly more relevant than irrelevant results, while a high recall means that an algorithm has yielded the most relevant results. The most amount of Tweets, Posts related with. Traffic, and because of this people can change their root and saving his/her time One of best thing in this project is we are using Stanford core NLP and C4.5classifier which was not present in existing system, Existing system uses the WEKA tool Means we can process very huge data within very short time period.

**VII RESULT AND DISCUSSION**

Key Index Parameters for Result Classification In information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also called sensitivity) is the fraction of the relevant instances that are retrieved. Precision and recall are therefore based on understanding and measuring relevance. In simple terms, high accuracy means that an algorithm returns significantly

more relevant than irrelevant results, while a high recall means that an algorithm has yielded the most relevant results, important category measurements for binary categories are:

**Precision (P)**  

$$P = TP / (TP + FP)$$

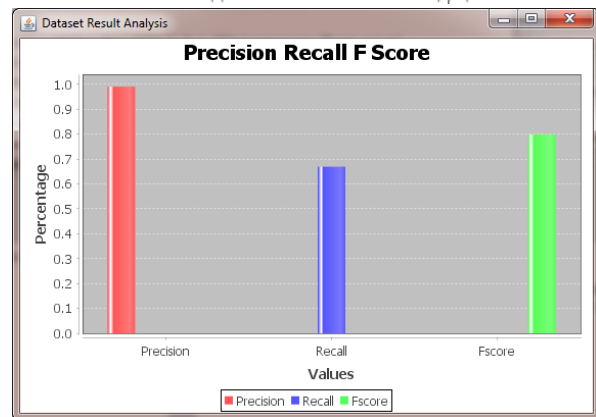
**Recall (R)**  

$$R = TP / (TP + FN)$$

**Accuracy (A)**  

$$A = \frac{tp + tn}{tp + tn + fp + fn}$$

$$Fmeasure = 2 * ((Precision * Recall) / (Precision + Recall))$$



**Figure 2 precision, recall Fscore**

**VIII ADVANTAGES**

The system solves the problem of traffic congestion in a cheapest and profitable way. It saves time by suggesting alternative routes. The system not only detects the traffic but also provides cause related to the traffic also suggested the alternative route. It is flexible and user friendly as it can be used anywhere or any devices that may be cell phones or PCs.

**IX APPLICATION**

This system can be used for tourism, Transportation, marketing and Advertisement .In the Transportation system its suggesting the suitable routes or transportation mean avoid the crowds, traffic jam or other problems that are related to the traffic.

**X CONCLUSION AND FUTURE SCOPE**

In this paper, we have proposed a system for dynamically analyzes the traffic by using twitter. The system is able to fetch and classify the tweets and notify users about the presence of traffic events and recommend alternative routes. The system can also provide the cause of the traffic efficiently. This system “is for mobile application is limited to android devices but in future it can be developed for iOS , Blackberry, Windows phone OS. The results of this work can be applied by enterprises in sentiment analysis to understand how their customers feel about a particular product or service and to track how those opinions change over time, and also to get

information regarding the relative performances of their competitors.[7].This provide Java with ,Stanford Core NLP attempt to do everything, Stanford Core NLP is used NLP toolkits This system uses the pipeline architecture. Stanford Core NLP [5]

#### REFERANCES

- [1] Dynamic Traffic Analyzer Using Twitter Harshita Rajwani<sup>1</sup>, Srushti Somvanshi<sup>2</sup>, Anuja Upadhye<sup>3</sup>, Rutuja Vaidya<sup>4</sup> , Trupti Dange RMD Sinhgad School of Engineering, Warje,Pune-58, Savitribai Phule Pune University
- [2] [2] Real-Time Sensing of Traffic Information in Twitter Messages Sara Carvalho Lu'is Sarmiento Rosaldo J. F. Rossetti, Member, IEEE
- [3] Detection of event from the twitter from twitter by Siqi Zhao.
- [4] ] P.Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620,
- [5] International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 02 Issue: 06 | Sep-2015
- [6] Real-Time Detection of Traffic From Twitter Stream Analysis Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member,.
- [7] The Stanford CoreNLP Natural Language Processing Toolkit Christopher D. Manning Linguistics & Computer Science Stanfordsmanning@stanford.edu Mihai Surdeanu SISTA University of Arizona.
- [8] The Movement in Real-Time Analysis Assistant Professor, Christu Jyothi Institute of Technology. Ch. Prudvini, Assistant Professor, ISSN No: 2348-4845 International Journal & Magazine of Engineering, Technology, Management and Research.
- [9] American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) ISSN (Print) 2313-4410, ISSN (Online) 2313-4402 © Global Society of Scientific Research and Researchers.
- [10] International journal of advance scientific research And engineering trends Live detection of traffic from twitter || Volume 1 || Issue 2 || May 2016 || ISSN (Online) 2456-0774