



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.569**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Data Mining based Innovative Method for Traffic Monitoring

V.Saranya<sup>1</sup>, P. Jeyakani<sup>2</sup>, Kanimozhi.M<sup>3</sup>, Naveen Kumar K<sup>4</sup>, Rameshkanth<sup>5</sup>

Department of Computer Science and Engineering, New Prince Shri Bhavani College of Engineering and Technology,  
Chennai, India<sup>1,3</sup>

Department of Information Technology, New Prince Shri Bhavani College of Engineering and Technology,  
Chennai, India<sup>2,4</sup>

Velsine Technologies Pvt. Ltd., Chennai, India<sup>5</sup>

**ABSTRACT:** Traffic congestion in lots of towns around the arena is severe. The cause is that society has in widespread become extra cellular and this indicates greater humans are prepared to go back and forth to paintings through automobile than they have been earlier than. To address these increasing congestion troubles we have give you positive strategies to factor out the site visitors congested areas. These techniques prove out to be useful to the pedestrians and travelers who regularly use roads as their method of transport. Monitoring traffic is a luxurious and time ingesting method this is why we've proposed a user pleasant proficiency to reveal social media updates which might be updated by means of the each day social media customers and helps as a tool for retaining the site visitors updates updated.

**KEYWORDS:** Traffic Analysis, Sentiment Analysis, Social Media Data, Data Science, Statistical Models.

## I. INTRODUCTION

The world populace is ever at the growth with the major populace dwelling within the city areas and metropolitan cities. The range of automobiles on the street is ever increasing and since the final five years it's growing at a fee of 10.Sixteen percentage yearly. Since the remaining five years, the quantity of vehicles on the road is growing. Traffic jams have become a part and parcel of life people living in these metropolitan cities. It no longer handiest results in wastage of fossil fuels which is a non-renewable supply of electricity however also wastage of time and money. Spending hours in traffic can also lead to fitness dangers and the burning of gasoline may cause environmental risks. There are some of methods or methods which are available or cautioned to address traffic issues. The primary approach is higher infrastructure to be built like wider roads, bypasses and expressways. But, for growing nations finance and availability of area to construct such infrastructure is probably a critical problem.

Another approach is the use of sensors and cameras. But these have issues like constrained insurance, highly-priced implementation and maintenance related to them. Another approach is probe-car facts. But these have problem like noisy facts on the arterials. Another approach advised makes use of Vehicle to Vehicle Communication, but those require deployment of equipments in vehicles and have worries like security and privateness associated with them. As those cities are technologically advanced they harbor big population of social media customers. Thus, we recommend a technique which makes use of assets that are readily to be had to those humans. Social media information can be leveraged to solve visitor's difficulty. As social media records are unstructured, we need to increase an interactive visual display which affords real insights into visitor's issues. These facts cannot best be used by the commonplace man but also respective visitor's authorities to take moves to slash traffic. We examine this social media facts to no longer handiest get site visitors-related facts however also sentiment patterns. We also can perform statistical analysis in this statistics to evaluate visitors among two cities.

## II. LITERATURE REVIEW

Traffic within the cities is divergent and is unpredictable. Road traffic congestion has constantly been a hard venture for the city areas. Moreover, monitoring such heavy site visitors the usage of sensors or probe cars is likewise no longer feasible because of the high installation cost and certain environmental situations [7]. The traditional method for tracking the visitors turned into by using the usage of sensors. Those sensing era additionally consisted of intrusive



and non-intrusive detectors. The non-intrusive sensors had been relatively skilled in sensing visitors speed, extent and classification as they protected infrared, excessive touchy, microwave radar, sound sensitive, ultrasonic and VIP sensors. Then a singular idea of monitoring the visitors the usage of social media which includes Twitter was brought up [8]. Twitter is an extensively used Social media networking web site. It is a rich source of information from across the globe. This social media is a great means of spreading information at earliest possible. Social media issued as an awareness component. The lengthy tail events can be detected due to sparsity [9][10]. Also monitoring of health events is performed. The device is skilled to test most effective for fireplace activities. Text classifier is used to test to check for related event of hobby and all of the associated events are grouped to form cluster. This record enables to alert the human beings dwelling in the vicinity. A text classifier is used to identify tweets of hobby which includes the ones accompanied with pix and tracking machine which could track multiple events without delay [11].

### III. RESEARCH METHODOLOGY

The information uploaded on social media via customers can be accessed the use of the REST API (Representational State Transfer Application Program Interface). Rest might be the most effective manner to use all the web offerings on information. It is fully-featured and is predicated on a stateless, customer-server, cacheable communicate protocol. It is not a trendy which lets you "roll your personal" basically make use of it in keeping with the way you would love to apply it with standard library functions in languages like Perl, Python, Java or C#. The Client makes a HTTP request to the REST Web Services Server. The server then establishes a streaming connection and problems a request to the Twitter REST API. Twitter accepts the relationship and streams Tweets as they arise. The HTTP Server pulls the data from the records keep and renders them into view. This is the response the Client receives. The patron can then see the tweets displayed on the internet site. The Twitter API additionally filters tweets with unique hashtags in step with the want of the particular utility. The REST makes use of HTTP for all of the 4 CRUD operations of Create, Read, Update and Delete. OAuth (Open Standard for Authorization) is used for comfortable access of records. It issues tokens which are to be issued to the customer by using an authorization server on behalf of the resource proprietor. Thus, this facts amassed the use of the Twitter REST API is used for in addition for analysis. The Figure beneath suggests server processes, one server receives the streamed tweets, even as the alternative brings the result and renders it into view.

Data Preprocessing is required as actual international information is incomplete, noisy and inconsistent .I.E. It lacks attributes of interest, contains mistakes and discrepancies in names and codes. Data Preprocessing is an effective step for evaluation of statistics. It removes undesirable, beside the point and noisy records and helps in giving appropriate great information.

The records this is acquired in the shape of tweets consists of records just like the message within the form of man or woman string, timestamp, location, consumer call, retweets, and many others. This records received is saved within the database and subjected to preprocessing which yields only those attributes which can be wished via the consumer or unique utility. This preprocessed statistics is once more saved into some other table of the database and analysis is achieved which includes removal of prevent words, stemming and calculating phrase frequencies. Sentiment analysis is finished in these records to achieve the precise meaning of the tweet and what the consumer is trying to express via his tweets on the social networking websites.

One of the main benefits of the usage of a Social Listening tool is that you can fast decide humans's emotions associated with any occasion, product. But Sentiment Analysis may be a subjective device, and those with some opinion are regularly located to figuring out the fantastic, negative or impartial sentiment from a comment or evaluation. Sentiment Analysis can be described as to routinely extract or classify sentiments from ordinarily unstructured data or textual content the usage of the aggregate of natural language processing (NLP) and the computational techniques. There are a few strategies which speeds up the processing of the analysis out of which Naïve Bayes is one of the efficient methods. It may be very easy to assemble as it does not require any complex iterative parameter estimation schemes. This specification in reality shows that it can be quite simply applied to massive datasets. It is easy to interpret, so it even becomes smooth to the customers to apprehend who are untrained in classifier technology can apprehend. And eventually, it regularly works quite properly. Though it can not be known as the best viable classifier, it is dependable in terms of it's performance and does pretty well.

The Google Maps API can be used for this cause. We can integrate the capabilities and offerings furnished through the Google Maps API to reap the routes and guidelines for travel along side the expected time to travel. As the tweets do no longer offer the range and longitude of locations particularly cities. We can maintain a database which



dynamically updates as locations are extracted from tweets. To take a look at if the place cited inside the tweets exists we will use Google Maps API for the equal purpose. The vicinity or town then received can obtain it is coordinates and keep those into the database. This information can then be used to plot maps as one of the foremost requirements at the same time as plotting the map are the coordinates of a place. One can then input the source and vacation spot of the vicinity they need to journey to and reap the riding directions and estimated time tour.

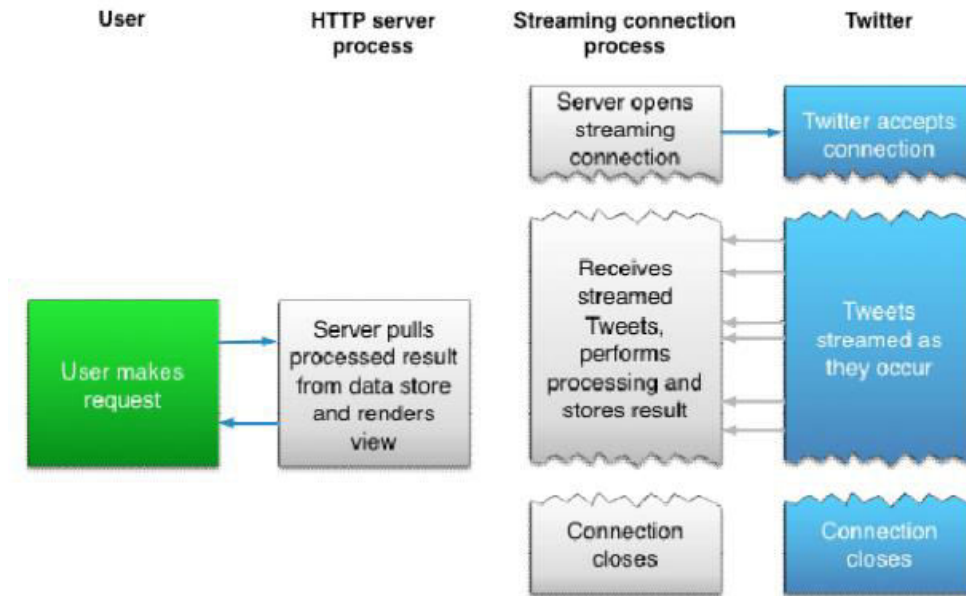


Figure.1 Flow and Processes between two Servers

#### IV. RESULTS AND DISCUSSIONS

We made an internet site in which we saved a login page, one from in which the users and the admin can login. Next, we created another page in which the tweets captured the usage of the Twitter REST API are displayed. We used a grid view with columns which display records like location, date and time stamp of the put up, person and the tweet message. We've enabled the paging in order that huge variety of tweets can be displayed. There's additionally a Import Live Tweets button at the internet site. On clicking the option the tweets are imported into the database. We've also made sure that most effective one copy of every tweet receives saved at the database and the duplicates are averted. This we made sure by if the tweets are imported into the database a number of times, simplest the tweets imported at the latest maximum click on are saved and the previous tweets of that day are discarded. At the end of the day most of these tweets are stored into the database as very last tweets for that day. Then, information extraction and cleansing is achieved on these tweets and the brand new listing of beneficial tweets is stored into another table of the database. The sentimental analysis is then accomplished on this beneficial listing of tweets and each tweet with it's message and sentimental analysis is displayed. Also, the pie charts showing sentiment analysis of traffic at the precise day of a specific metropolis is displayed.

This is simply one of the methods you may design your interactive Dashboard according to the way you want it to appearance to increase aesthetics. Important is the way you carry out the evaluation. We additionally created every other choice on our Dashboard where you can use the maps for visiting directions. We used the Google Maps API functionalities and included it into our internet site. So now the consumer cannot simplest see the path from it is direction to vacation spot and get the estimate of the journey time however also can keep away from routes which might be crowded or have bad sentiments associated with them on the Dashboard and can take an alternate route. Figure underneath indicates the maps and the direction from Pune to Mumbai and the estimate of travel time on the driving course panel. The efficiency of our machine will increase as time passes and increasingly database or information is acquired as the complete method is primarily based not only on the quantity of records but additionally the first-rate of statistics. As the machine receives skilled on increasingly statistics and comes across extra real-time consequences after their analysis the performance of the system increases and after a point of time it's far envisioned to carry out extra correct than ever.



## V. CONCLUSIONS

This work addresses the problem of traffic congestion via specializing in monitoring of the traffic using Twitter as the handy social media. As the result of site visitors monitoring goes to guide the general public, the processing need to be completed with tons more accuracy. So the related works commonly awareness on big-scale occasion detection. Large quantity of statistics associated with any specific visitors related occasion leads to greater unique idea approximately the event passed off, and for this reason enables with correct outcomes. In our paintings, we exhibit how a language model and Naive Bayes classifiers may be combined to identify visitors congestion places and show the facts of sentiment evaluation the use of a pie-chart. The machine is able to retrieve and classify streams of tweets which obviously offer an alert signal to the customers concerning the lifestyles of traffic occasions.

## REFERENCES

1. Sentiment Classifier in PHP, <https://github.com/JWHennessey/phpInsight>.
2. Twitter API Wrapper for Twitter API v1.1 Calls, <https://github.com/J7mbo/twitter-api-php>.
3. FacebookGraphAPIExplorer, <https://developers.facebook.com/tools/explorer>.
4. Bo Pang, Lillian Lee, Opinion mining and sentiment analysis, 2008.
5. NaiveBayesClassifier, [http://en.wikipedia.org/wiki/Naive Bayes classifier](http://en.wikipedia.org/wiki/Naive_Bayes_classifier).
6. Google Maps JavaScript Library, <https://github.com/hpneo/gmaps>.
7. Yu, P. D. Prevedouros, Performance and Challenges in Utilizing Non Intrusive Sensors for Traffic Data Collection, *Advances in Remote Sensing*, 2013, 2, 4550.
8. Jianshu Weng; Yuxia Yao; Erwin Leonardi; Francis Lee, HP Laboratories HPL201198 Event Detection in Twitter. In ICWSM, 2011.
9. P. Agarwal, R. Vaithyanathan, S. Sharma, G. Shroff. Catching the Long-Tail: Extracting Local News Events from Twitter. In ICWSM, 2012.
10. Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, Hongjun Lu. Parameter free bursty events detection in text streams. In VLDB, pp. 181192, 2005.
11. Aiello, L. M., Petkos, G., Martin,. Sensing trending topics in Twitter , 2013.



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details