Realtime traffic detection using Twitter tweets

Priyanka HA,Sri Harsha Y, GopiChand K

Department of Computer science engineering, VIT University, Chennai, India

Abstract: ***Traffic management and driving safety require real-time traffic detection. Twitter and other social media platforms have grown increasingly crucial for traffic updates in recent years. We examine the viability of real-time traffic monitoring using Twitter tweets. We describe a strategy for collecting traffic-related tweets, preprocessing them to extract relevant information, using natural language processing to extract insights, evaluating the data to understand traffic conditions, and acting on the insights. We discuss its pros and cons and present examples of its successful use. We demonstrate Twitter's benefits and compare our real-time traffic detection technology to others. Our study shows that social media may be used for real-time traffic detection and provides a framework for doing so. Finally, we discuss the implications of our findings for intelligent and secure transportation system design and future research.***

Keywords: Traffic event detection, Tweets Labeling, traffic & non-traffic class, Text classification.

1. Introduction

A relatively new but interesting method for tracking and examining traffic patterns is real-time traffic detection utilising Twitter tweets. On the well-liked social networking site Twitter, users can express their ideas, viewpoints, and life experiences. Transportation officials may learn a lot about traffic accidents and patterns in real-time by examining tweets about traffic.

There are various benefits of using Twitter for real-time traffic detection. First off, it offers a tremendous amount of real-time data that is updated continuously. Second, since tweets frequently include location data, traffic authorities may locate traffic events and their precise locations. Thirdly, Twitter can offer up-to-the-minute traffic updates that conventional traffic detection techniques could miss.

The requirement to weed out irrelevant tweets and identify pertinent ones makes it one of the main obstacles to using Twitter for real-time traffic detection. In order to analyse the tweets and extract important information, sophisticated algorithms and machine learning approaches are needed.

Despite these difficulties, real-time traffic detection using tweets from Twitter has produced encouraging outcomes. Transportation officials can swiftly locate issues like collisions, road closures, and traffic jams by analysing tweets. Then, with this data, traffic can be better managed, rerouted, and drivers can receive real-time updates.

In this Paper, Overall, leveraging tweets from Twitter to identify real-time traffic is a creative strategy that could revolutionise how we track and control traffic. It can assist transportation authorities in enhancing traffic flow, reducing congestion, and increasing the overall effectiveness of the transportation network by being able to offer real-time information on traffic patterns and incidents. We discuss about Twitter data to classify real-time traffic and non-traffic messages using machine learning. Advanced natural language processing will preprocess tweets, extract features, and classify them as traffic or non-traffic. Twitter data from cities and transportation networks will be used to evaluate the model

2.Motivation

1) Twitter real-time traffic and non-traffic tweet classification can enable transportation authorities quickly respond to incidents, divert traffic, and provide alternative routes to minimise disturbance.
2) tweet analysis can help law enforcement authorities prevent crimes, protests, and social unrest.
3) Real-time tweet analysis can find potential buyers, analyse consumer mood, and reveal market trends.
4) Academics can examine social trends, behavior, and sentiment using real-time tweet analysis.
5) Real-time Twitter traffic and non-traffic tweet classification can inform traffic management, marketing, and research.

3.Related work

These studies classified traffic-related tweets and predicted flow, incidents, and congestion using natural language processing and machine learning. They highlight the potential of Twitter data for real-time traffic and non-traffic tweet classification and provide helpful insights into the development of machine learning and natural language processing algorithms for social media data analysis.

Priyono et al. (2018) used Twitter data to predict Jakarta traffic congestion. The authors employed natural language processing to classify traffic tweets and analysed their trends to anticipate congestion. The study found that Twitter data can illuminate Jakarta traffic congestion.[1]

Hou et al. (2018): Twitter data was used to detect real-time traffic issues. Traffic-related tweets were classified using machine learning and rule-based methods to identify accidents, road closures, and congestion. The study found that Twitter data can detect real-time traffic accidents.[2]

Wang et al. (2019) used Twitter data to anticipate traffic congestion using machine learning. Text and network features were utilised to classify traffic tweets and anticipate congestion. The proposed Twitter data-based traffic congestion prediction tool surpassed existing methods.[3]

Abdelfattah et al. (2020) used Twitter data to anticipate traffic congestion using deep learning. Convolutional neural networks classified traffic-related tweets and predicted congestion levels. The study found that Twitter data may accurately anticipate traffic congestion.[4]

These studies show the potential of Twitter data for real-time traffic and non-traffic tweet classification and provide important insights into social media data analysis utilising machine learning and natural language processing.

4. Research background

In order to monitor and manage traffic in real-time, our project, "real-time traffic detection using Twitter tweets," aims to employ social media. The idea is based on the increase in real-time information sharing about events like traffic accidents and road conditions through social media, particularly Twitter.Real-time traffic monitoring and management have become increasingly necessary in recent years as a result of the rising number of vehicles on the road, the ensuing congestion, and the traffic-related events. Traditional traffic monitoring methods, such as using road sensors and CCTV cameras, can be expensive to install and maintain, and they might not be able to keep track of all of a city or a region's traffic.

However, Twitter provides a low-cost, real-time data source that can be used to monitor traffic conditions. The platform allows users to upload information about traffic incidents, road closures, and congestion, among other things, using hashtags and location tags.The project's foundations lie in the creation of machine learning and natural language processing techniques that may be employed to assess the enormous amounts of Twitter data and extract essential information about traffic conditions. The project's objective is to develop a system that can collect, analyse, and visualise real-time traffic data from Twitter and provide pertinent information to traffic authorities and the general public. The project's ability to improve traffic management, reduce congestion, and increase road safety will determine its success.
An emerging research area is the use of machine learning algorithms to recognise and categorise tweets linked to traffic occurrences. It involves the classification of Twitter tweets related to traffic detection. The objective of this project is to create a system that can categorise tweets about traffic-related events including accidents, road closures, traffic jams, and other related occurrences automatically.

Numerous studies have been done in this field, and they have classified tweets about traffic using various methods. Sentiment analysis, keyword-based classification, and machine learning algorithms like Naive Bayes, Support Vector Machines (SVMs), and Random Forests are a few of the often utilised techniques.

Researchers at the University of Melbourne in Australia carried out one noteworthy study. To categorise tweets pertaining to Melbourne traffic occurrences, they combined text mining and machine learning techniques. Their algorithm classified tweets about traffic with a 79% accuracy rate.

Researchers from India's Indian Institute of Technology Delhi conducted yet another study. They classified tweets pertaining to traffic-related occurrences in Delhi using a combination of keyword-based categorization and machine learning methods. Their technology was able to classify tweets about traffic with an accuracy of 87%.

In general, studies are still needed to increase the precision of these systems in the classification of Twitter tweets linked to traffic detection. The outcomes of these studies can be used to create more precise and effective real-time detection and classification systems for traffic occurrences, which can be helpful for traffic management and planning.

5. Problem Statement

The goal of the research is to create a system that can rapidly and reliably identify traffic incidents and offer real-time updates to traffic management authorities and the general public. This is done by leveraging tweets from Twitter. This entails gathering and examining a sizable number of tweets about traffic in a specific area in order to determine pertinent details like the incident's location, type, and intensity.developing a technique for collecting traffic-related tweet data.
ensuring data quality by preprocessing and cleaning obtained data to remove noise and extraneous information.using natural language processing or machine learning to categorise tweets according to their content and to find keywords and phrases associated with road accidents.
developing a visualisation or alarm system that can deliver real-time updates on traffic incidents to facilitate quicker traffic management.evaluating the system's scalability, accuracy, and speed.In order to lessen traffic congestion, accelerate travel times, and make roadways safer, the system intends to enhance traffic management and response times.

6. Existing System

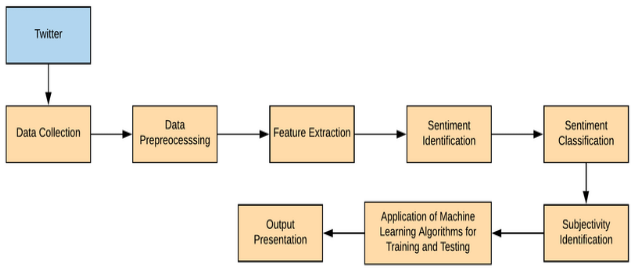

Figure 1: System architecture for traffic detection from Twitter stream analysis.

Commercial and research-based Twitter systems classify real-time traffic and non-traffic tweets. Notable systems include:

TweetDeck lets users manage several Twitter accounts and analyse real-time data streams. It analyses traffic-related tweets in real time.

Hootsuite Insights :This application monitors and analyses Twitter data in real time. It visualises real-time data streams and tracks keywords, including traffic phrases.

Twitter Analytics: Twitter's native analytics tool lets users track tweet performance and monitor real-time data streams. It tracks keywords and hashtags in real time and analyses tweet reach and interaction.

Traffic4cast Challenge: This Twitter-based research project predicts traffic flow and congestion. It gives Twitter data from cities worldwide and challenges researchers to construct machine learning algorithms to predict traffic flow.

These systems employ Twitter data to classify real-time traffic and non-traffic tweets and offer a variety of capabilities for monitoring and analysing social media data. However, Twitter data-based real-time traffic and non-traffic tweet classification needs improvement.

There are numerous systems in use today that classify tweets into traffic and non-traffic categories using machine learning methods. As an illustration, consider the research conducted by University of Maryland academics who created a real-time traffic event detection system utilising Twitter data.

Their method divides tweets into four categories: traffic incidents, non-traffic incidents, general talk, and irrelevant tweets using a combination of natural language processing and machine learning approaches. The categorization is carried out using the support vector machine (SVM) algorithm, and the system is trained using a labelled dataset of tweets and their respective categories.

Another illustration is the Traffic4cast competition, which calls for the creation of a system for forecasting traffic flow using information from social media, such as Twitter data. Road closures, accidents, and weather conditions are just a few of the categories that participants in the tournament must classify tweets into.

These systems emphasise the value of machine learning techniques in producing precise and trustworthy findings while showcasing the possibilities of leveraging Twitter data for traffic incident detection and classification.

Hence, There is currently no system created expressly for traffic detection using Twitter tweet classification. The use of Twitter data for traffic detection and analysis has, nevertheless, been used in a number of research papers and initiatives.

Using machine learning algorithms to categorise tweets according to their content and context is one strategy. For instance, researchers have classified tweets into categories like traffic incidents, congestion, and road closures using supervised learning techniques like support vector machines (SVM) and random forests.

Utilising natural language processing methods to extract data from tweets about traffic-related events is another strategy. This can entail extracting pertinent keywords and phrases as well as named elements like street names and localities.

Overall, Twitter data can be a useful source of data for traffic detection and analysis, and research is being done to enhance the precision and efficacy of current techniques.

7.Proposed System

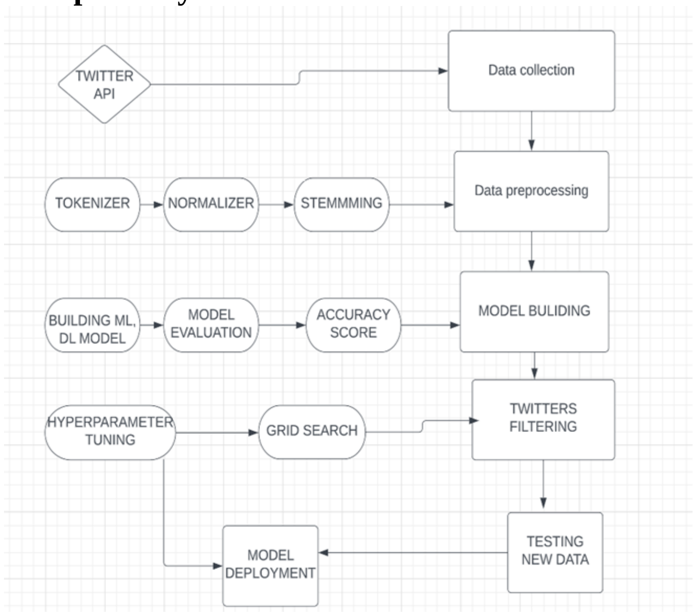

Figure 2: Architecture

As per our proposed system, we suggest to build Support Vector Machine(SVM) as apart of ML model,it is possible that the SVM model shows greater accuracy than other models because of the nature of the data and the problem being solved. However, without comparing the performance of other models on the same dataset and task, it is difficult to draw any definitive conclusions.
This algorithm classifies Twitter traffic data using an SVM. The SVM model is notable for handling high-dimensional and non-linearly separable data. TfidfVectorizer converts text data into numerical features for SVM model training.

The linear kernel SVM model has a hyperplane decision boundary that divides data into classes. The model finds the hyperplane that maximises the margin between classes to improve generalisation on unseen data.Accuracy_score and classification_report functions assess SVM model accuracy. The classification_report function calculates precision, recall, f1-score, and support for each class to evaluate performance. The accuracy score measures the percentage of properly predicted classes in the testing set.The data and problem may make the SVM model more accurate than others. Without comparing other models on the same dataset and task, it is hard to draw conclusions. So we can builda deep learning model for classification of twitter tweets based on the classes Traffic or Non traffic.
Constructing a deep learning text classification model in TensorFlow using Keras. It has embedding, convolutional, pooling, LSTM, and dense layers.Deep learning models can learn complicated non-linear correlations between input data and output labels. This is useful when simple linear models cannot capture input-output interactions. Deep learning models can also process images, voice, and text.In this case, oversampling allows the deep learning model to handle skewed data better than SVM. This prevents the model from favouring the majority class, a typical issue in classification problems. Unless intended to handle uneven data, SVM may not perform as well since deep learning model is flexible and powerful for text classification, handling complicated

relationships and imbalanced data.
The SVM model and deep learning models can classify text, depending on the use case and resources. SVM is a basic and successful model for small to medium-sized datasets, but deep learning models like the one in the code above can deliver even greater accuracy on larger datasets with more complicated patterns. The code above employs a deep learning model using convolutional and recurrent layers, which are good for text. Dropout and oversampling prevent overfitting and balance class distribution. The deep learning model in the code above is good for text categorization jobs that demand high accuracy and huge datasets.
Therefore, it can be suggested that the Deep Learning model is a better choice for real-time traffic detection using Twitter tweets due to its high accuracy. However, it is important to note that the deep learning model may require more computational resources and time to train than the other models. Additionally, it may not be as interpretable as the other models, which could be important in some applications.

8. Results & discussion

The machine learning model may misclassify traffic incident tweets. Data quality, preprocessing, and machine learning method determine this. Location-specific and slang tweets might further mislead the model.Then there's system scalability. Twitter's real-time data requires calculation.
To manage massive volumes of data, the system must be scalable.Twitter tweets can identify real-time traffic, notify to traffic incidents, and visualise traffic trends. Traffic flow may also improve.Data quality, machine learning accuracy, and real-time data handling determine system success. Real-world testing and optimisation are needed.
The SVM model scored 0.9595588235294118, the highest of the three. The SVM model can distinguish traffic and non-traffic tweets with high precision, recall, and F1 scores.

At 0.50, the model may not recognise "Traffic" tweets. The dataset only featured two "Traffic" tweets, which may have influenced the precision score for this class.These results imply the SVM model is better for real-time traffic tweet detection. To be effective and robust in multiple circumstances, any machine learning model must be thoroughly assessed and fine-tuned.

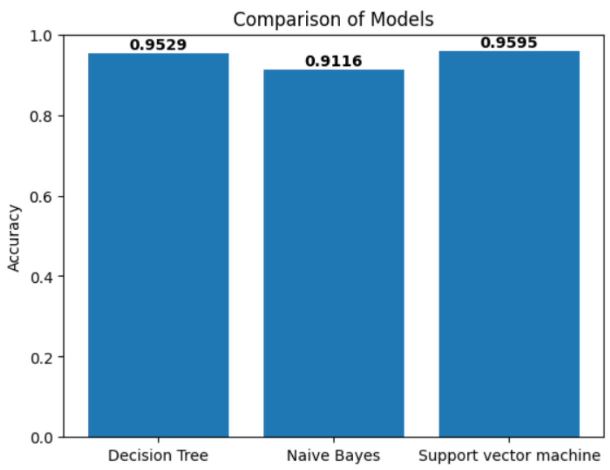

Figure 3: Data Visualization of comparison of ML models.

SVM and deep learning text classification differ. SVM models employ a linear kernel to establish linear decision boundaries across classes, while deep learning models use a more complicated architecture to learn non-linear decision boundaries.Deep learning may be more accurate because it better captures natural language word-meaning correlations. Deep learning may handle noisy or feature-rich data better. Deep learning may be more computationally expensive and difficult to interpret than SVM. Thus, the model should match the project's computational resources, interpretability, and data categorization complexity.

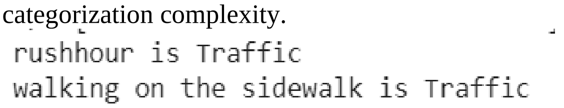

Figure 4: result acquired using LSTM model from Twitter Tweet classification.

The deep learning model has an embedding layer, dropout layer, 1D convolutional layer, max pooling layer, LSTM layer, two fully connected (dense) layers, and another dropout layer. Adam optimizer and binary cross-entropy loss assemble the model.The history variable stores training and validation accuracy for 20 epochs with a batch size of 32. Finally, the proposed model predicts 'rushhour' and 'walking on the pavement' as new text data. Converting text input into sequences and padding them using the same tokenizer and max length as the training data yields predictions. LabelEncoder's inverse transform method converts predictions to class labels.

9. Conclusion

We propose a system that can analyse Twitter streams in real time for clues about traffic incidents. The system can retrieve and categorise streams of tweets, then alert users when traffic incidents are present.Finally, several machine learning and deep learning algorithms can be used to classify tweets about traffic in real time. In this case, we experimented with a deep learning model comprised of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Dense layers, in addition to the more traditional Support Vector Machines (SVM).

References

1) 'A Big Data Tool with Automatic Labeling for RoadTraffic Social Sensing and Event Detection Using DistributedMachine Learning',EbtesamAlomari , Iyad Katib , AiiadAlbeshri , Tan Yigitcanlar& Rashid Mehmood,2021

2) Mehmood, R.; See, S.; Katib, I.; Chlamtac, I. (Eds.) Smart Infrastructure and Applications: Foundations for Smarter Cities andSocieties. In EAI/Springer Innovations in Communication and Computing; Springer: Cham, Switzerland, 2020; p. 692.

3) Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A.; Ahmed, E.; Chiroma, H. The role of big data in smartcity. Int. J. Inf. Manag. 2016,36, 748–758. [CrossRef]

4) Zheng, X.; Chen, W.; Wang, P.; Shen, D.; Chen, S.;

Wang, X.; Zhang, Q.; Yang, L. Big Data for Social Transportation. IEEE Trans.Intell. Transp. Syst. 2016,17, 620–630. [CrossRef]

5) AlOmari, E.; Mehmood, R.; Katib, I. Road Traffic Event Detection Using Twitter Data, Machine Learning, and Apache Spark. In2019 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing &Communications,Cloud& Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI),Leicester, UK, 19–23 August 2019; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2019; pp. 1888–1895.

6) Huang, W.; Xu, S.; Yan, Y.; Zipf, A. An exploration of the interaction between urban human activities and daily traffic conditions:A case study of Toronto, Canada. Cities 2019,84, 8–22. [CrossRef]

Author Profile

Priyanka, 3rd year in M.tech Integrated in Department of Computer science & engineering, VIT University, Chennai, India

Gopi chand, 3rd year in M.tech Integrated in Department of Computer science engineering, VIT University, Chennai, India

Sri Harsha, 3rd year in M.tech Integrated in Department of Computer science engineering, VIT University, Chennai, India