

Prediction of Traffic Congestion Utilising Machine Learning and Big Data Analysis for Smart Cities

Prashant Kumar^{1,a)}, Prof. Dr Renu Bagoria^{2,b)}, Mr Suraj Yadav^{2,c)}, Pranab Kumar Saha^{3,d)} Deepika kumari^{3,e)}

1) Department of Engineering (Computer Science), Jagannath University, Jaipur, Rajasthan, India

2) Department of Engineering (Computer Science), Jagannath University, Jaipur, Rajasthan, India

Corresponding authors: a) renu.bagoria@jagannathuniversity.org

b) suraj@jagannathuniversity.org

c) prasahantsinghmathacademy@gmail.com

d) pranabkumars52@gmail.com

e) deepika.kumari08913@gmail.com

Abstract—Rapid urbanisation and urban population growth, combined with the growing number of vehicles, have made traffic congestion among the biggest challenges of modern cities. Traffic congestion increases travel time, fuel consumption, pollution, and economic losses. In smart cities, there is a large volume of transportation data generated from heterogeneous sources such as GPS trackers, traffic sensors, CCTV cameras, weather reports, and mobile applications. Intelligent analysis of such huge volumes of transportation data will allow conducting effective traffic management in modern cities.

In this research, a Predicting Traffic Congestion Using Machine Learning and Big Data Analysis Framework for Smart Cities system was suggested that aims at analysing historical and live data on traffic and predicting the congestion of roads prior to the appearance of severe traffic problems. The proposed solution involves conducting data preprocessing, feature engineering, and predictive modeling processes that allow recognising patterns in the data and predicting possible states of congestion. The following machine learning algorithms were used in order to improve the results of prediction: Random Forest, Gradient Boosting, and LSTM neural network models.

Keywords— Traffic Congestion Prediction, Machine Learning, Big Data Analytics, Smart Cities, Intelligent Transportation Systems, Predictive Analytics, Traffic Forecasting, Data Science

1. Introduction

Traffic congestion has become one of the most vital issues that urban areas face due to their rapid development. The constant increase in population and cars, urbanisation, as well as the lack of transport facilities, lead to significant traffic density, causing serious issues related to traffic congestion. Traffic congestion affects not only travel times but also fuel consumption, air pollution, road accidents, and losses. The traditional traffic management system uses static methods and fails to effectively cope with ever-changing traffic situations in real-time. The use of smart city innovations leads to the creation of massive volumes of transportation data through various sources such as GPS, surveillance cameras, road sensors, apps, weather observation tools, and IoT devices. Massive and continuous data generation allows the use of advanced techniques in order to enhance traffic management and mobility in cities. Big Data Analytics combined with Machine Learning becomes an efficient way of processing such data and obtaining accurate results in terms of traffic prediction. Machine learning algorithms allow detecting underlying relations in traffic data sets. Such algorithms as Random Forest, Decision Trees, Gradient Boosting, and Deep Learning provide high-level traffic analysis and accurate traffic congestion prediction. The research will develop a model on Prediction of Traffic Congestion using Machine Learning and Big Data Analysis in Smart Cities. The main objective of the system is to evaluate the traffic condition and predict the level of traffic congestion by considering various traffic parameters such as number of vehicles, speed of movement, road occupancy, weather conditions, time intervals, among others. The predictions will be helpful in decision-making related to traffic management, route optimisation, and infrastructure planning. Some of the key

purposes of conducting this research include accurate prediction, reduction of delays, improving transportation efficiency, and promoting sustainability. The contribution of the research in relation to ITS is evident since it will enhance traffic management.

2. Literature Review

Urban traffic congestion has become a major issue in modern times owing to an increasing rise in population and vehicle numbers. Traditional urban traffic control models used fixed-time traffic lights and rule-based systems for traffic management. Even though those systems could manage traffic, they were unable to adapt to changes in real-time. Data-driven techniques have since been suggested for better prediction and intelligent urban traffic planning by researchers. In early studies, researchers used statistical methods to forecast the traffic flow and predict traffic volumes by analysing historical traffic datasets. Techniques such as linear regressions, time-series modelling, and probabilistic modelling were adopted to make predictions. Despite being able to model traffic, the problem with these methodologies was their inability to consider the non-linear aspects of the system. Since there was an emergence of sensor networks and digital transportation systems, huge amounts of data are generated from GPS technology, cameras, and other surveillance devices installed. The development of such technologies led to the adoption of prediction tools to enhance the accuracy of prediction models.

The use of Machine Learning techniques has emerged as a popular approach in traffic congestion prediction due to their ability to discover implicit patterns through big data analysis. Several studies explored different types of supervised learning algorithms for traffic condition predictions, such as Decision Trees, Random Forest, Support Vector Machines, Gradient Boosting, and Artificial Neural Networks. It has been found that Random Forest provides an effective prediction tool for traffic conditions due to its ability to process huge data and avoid overfitting with the help of an ensembling technique. The models based on Support Vector Machines have also proven their effectiveness at classification-based congestion prediction, particularly dealing with nonlinear traffic relationships. More advanced solutions based on deep learning, such as LSTM and RNN, were developed to predict the future traffic state by taking into account dependencies within sequences. They provide better accuracy results compared to other algorithms. However, there is still a problem in processing multidimensional traffic information and increasing computation efficiency in predicting processes in real-time. Existing literature focuses on hybrid models combining several algorithms for effective intelligent transportation.

The development of smart city infrastructure revolutionised traffic management through the use of consistent data generation and the application of big transportation analytics. Modern cities produce huge amounts of structured and unstructured traffic information through the Internet of Things devices, GPS systems, surveillance networks, weather monitoring systems, and applications on mobile phones. However, traditional methods of data processing may prove to be ineffective at dealing with such enormous volumes of data. Big Data Analytics plays an important role in intelligent transport systems by allowing for distributed processing and real-time decision-making based on the results of big data analysis. Technologies associated with big data analytics help researchers process traffic data and generate reliable congestion predictions. According to available literature, the combination of real-time data and historical datasets produces more accurate traffic predictions and helps detect emerging congestion patterns in advance. In addition to that, recent research has explored the use of cloud computing architecture and edge computing systems for the purpose of decreasing traffic monitoring delays. In smart cities, traffic prediction techniques play an important part in transportation planning and optimising the distribution of routes to minimise fuel consumption.

Despite considerable successes already made by traffic congestion prediction researchers, there are still certain shortcomings associated with current practices in this area. First, many studies are highly reliant on historical data, while they ignore other critical factors like traffic dynamics, weather conditions, road accidents, and other similar phenomena. In some cases, accurate predictions are provided; however, the problem is that they take too much time to compute. Interpretability is another shortcoming identified in recent literature in relation to advanced ML and DL prediction algorithms applied to traffic management. Since transportation agencies need to have full visibility in what happens on the roads and how traffic behaves at certain times, they prefer interpretable approaches. Also, some algorithms are created based on regionally limited datasets. In order to fill this gap, the following research will develop a new traffic

congestion prediction solution relying on big data analytics and machine learning technologies. Such a system will be able to make use of all the available information about traffic and provide reliable predictions to help manage traffic better.

3. Methodology

3.1 Overview of Proposed Methodology

This study introduces an innovative approach to predict traffic congestion that combines the application of Machine Learning algorithms along with Big Data analytics in the context of a smart city. The proposed technique can gather a massive amount of transportation data from various sources, analyse the information, create predictive models, and

produce real-time predictions on traffic congestion. The purpose of the proposed methodology is to design a scalable and accurate prediction system that can be applied in the context of intelligent transportation systems.

3.2. Data Collection

Data collection is the initial and most important step of the system. Traffic data is collected from various types of sources to provide an accurate representation of the traffic situation.

Sources of Data:

- a) GPS Tracking System
- b) Closed-Circuit Television Cameras (CCTVs)
- c) Road Sensors
- d) Mobile Navigation Systems
- e) Weather Stations
- f) Historical Databases
- g) IoT Devices in Smart Cities

The acquired data set comprises structured and semi-structured data to cover all possible traffic-related parameters.

- a) Data Set Attributes:
- b) Number of Vehicles
- c) Average Speed of Vehicles
- d) Timestamp
- e) Occupancy of the Road
- f) Temperature
- g) Rainfall Situation
- h) Traffic Light Time
- i) Road Capacity
- j) Coordinates of the Location
- k) Congestion Level

The collected data set will be stored at a central location for further processing purposes.

3.3 Data Preprocessing

Incomplete data, duplicate entries, noise, and inconsistent data present in raw traffic datasets can impair the effectiveness of models. Preprocessing should be done prior to modelling.

Preprocessing Tasks:

Data Cleaning- The tasks performed include:

- a) Elimination of duplicates
- b) Addressing missing data
- c) Detection of outliers
- d) Noise filtering

Data Transformation - Transformation is the process where data collected is converted into machine-readable data. The tasks performed include:

- a) Normalisation
- b) Standardisation
- c) Categorical data encoding

3.4 Feature Engineering and Feature Selection: The process of feature engineering enhances the predictive power of the machine learning model by deriving valuable features from the data set.

Derived features:

- a) Presence Indicator for Peak hour
- b) Number of Vehicles
- c) Average time for travel
- d) Weather Severity Factor
- e) Congestion Index (Past)
- f) Speed variation factor

Techniques are used for selecting valuable features while discarding unnecessary ones, reducing the computational costs. Techniques include:

- a) Correlation Study
- b) Feature Importance Score
- c) Recursive Feature Elimination

3.5 System Architecture

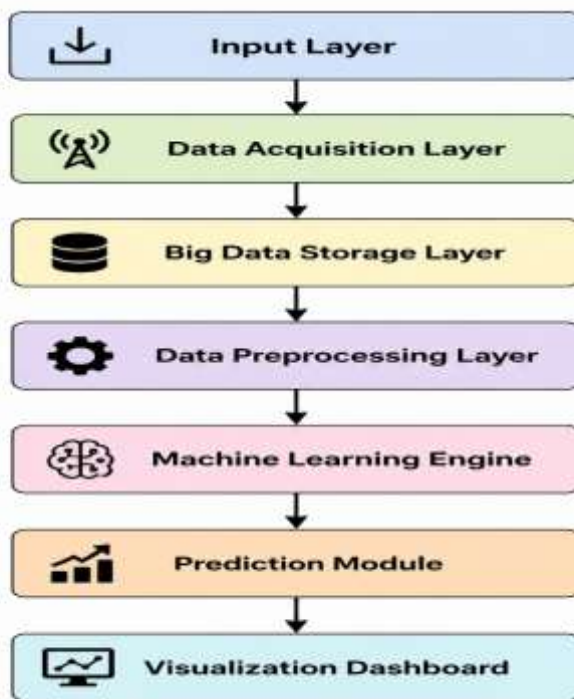


Fig 1.1 System Architecture

4. Future Scope

- a) Some potential improvements of the suggested Traffic Congestion Prediction System include:
- b) **AI and Deep Learning Methods Implementation-** Further development of the suggested prediction system can leverage state-of-the-art AI algorithms like LSTM, GRU, and transformers for even more precise prediction.
- c) **Integration with Real-Time IoT Sensors** -This framework may connect to IoT-based sensors that provide real-time traffic data from traffic lights, CCTV cameras, GPS devices, and cars.
- d) **Application in Smart Cities-** The algorithm may prove to be useful for smart cities and intelligent traffic signals and routing systems.
- e) **Using Edge Computing along with Cloud Platforms** -The proposed model might be implemented in conjunction with both edge computing and cloud infrastructure for faster traffic processing.
- f) **Incorporating Additional Sources of Information-** In addition to traffic data, other types of data, including weather information, traffic accidents, road construction, public events, and social media updates, might be leveraged for better predictions.

5. Conclusion

Urban environments face traffic congestion today because of the increasing number of vehicles due to the rapid rise of the urban population and increased demands for transportation services. In this regard, the paper described a Traffic Congestion Prediction System (TCPS) that used Big Data Analytics to make efficient predictions by analysing large transportation data volumes.

The proposed model involves different steps, such as acquiring, storing, preprocessing data, predicting, and visualising. The idea is to implement several algorithms for analysing the obtained information in order to improve prediction accuracy and optimise the process of decision-making related to traffic. Thus, the application of Big Data technologies and predictive analytics can provide significant advantages in terms of reducing congestion, selecting effective routes, lowering fuel consumption, and making travel more efficient. The proposed model is scalable and adaptive, and it may be applied for further implementations in smart cities.

Traffic congestion has emerged as one of the biggest threats for modern urban societies owing to factors such as rapid urbanisation, increase in populations, and high densities of vehicles. Poor traffic management is associated with various problems including time wastage, more fuel usage, pollution, and economic inefficiencies. As such, the need for innovative, intelligent and analytical traffic management systems has emerged for sustainable urban living.

This study proposed a Traffic Congestion Prediction System (TCPS) for effective traffic management using Big Data Analytics and transportation data sets. The TCPS framework consists of various layers such as acquisition of transportation data, storage, preprocessing, prediction, and visualisation modules to enable efficient decision-making through real-time analytics.

It was established in the study that a combination of Big Data technologies and analytical processes allows for the identification of various patterns in traffic data obtained from different sources such as GPS technology, traffic sensors, CCTV monitoring systems, and mobile technologies. The use of such an approach will facilitate better traffic management.

6. References

- [1] S. Rosenbloom, **"Peak-period traffic congestion: A state-of-the-art analysis and evaluation of effective solutions,"** Transportation, vol. 7, no. 2, pp. 167–191, 1978.
- [2] A. Downs, **Stuck in traffic: Coping with peak-hour traffic congestion.** Brookings Institution Press, 2000.
- [3] T. Litman, **"Transportation cost and benefit analysis,"** Victoria Transport Policy Institute, vol. 31, 2009.
- [4] A. Akbar, A. Khan, F. Carrez, and K. Moessner, **"Predictive analytics for complex iot data streams,"** IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1571–1582, 2017.
- [5] M. Deshpande and P. Bajaj, **"Performance improvement of traffic flow prediction model using combination of support vector machine and rough set,"** International Journal of Computer Applications, vol. 163, no. 2, pp. 31–35, 2017.
- [6] W. Min, L. Wynter, and Y. Amemiya, **"Road traffic prediction with spatio-temporal correlations,"** in Proceedings of the Sixth Triennial Symposium on Transportation Analysis, Phuket Island, Thailand (June 2007), vol. 65, p. 85, 2007.
- [7] Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang, et al., **"Traffic flow prediction with big data: A deep learning approach,"** IEEE Trans. Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, 2015.
- [8] J. Lee, B. Hong, K. Lee, and Y.-J. Jang, **"A prediction model of traffic congestion using weather data,"** in 2015 IEEE International Conference on Data Science and Data Intensive Systems, pp. 81–88, IEEE, 2015.
- [9] J. Ahn, E. Ko, and E. Y. Kim, **"Highway traffic flow prediction using support vector regression and Bayesian classifier,"** in 2016 International Conference on Big Data and Smart Computing (BigComp), pp. 239–244, IEEE, 2016.
- [10] D. Basak, S. Pal, and D. C. Patranabis, **"Support vector regression,"** Neural Information Processing-Letters and Reviews, vol. 11, no. 10, pp. 203–224, 2007.