International Journal of Multidisciplinary Engineering in Current Research - IJMEC

Volume 10, Issue 6, June-2025, http://ijmec.com/, ISSN: 2456-4265

Integrating Machine Learning Into Intelligent Transportation Systems For Real-Time Traffic Prediction

Mohammed Abdul Muqtadir [Phd]

Lecturer, Mazoon College, Affilated to Missourie University of Science and Technology

mabdulmuqtadir@gmail.com

Abstract: Despite the numerous safety features that automakers have created to reduce the likelihood of traffic accidents, accidents nevertheless happen regularly in both urban and rural locations. The development of precise prediction models that can recognize patterns connected to various events is essential to avert mishaps and enhance safety protocols. We may group accident possibilities and create efficient safety precautions by employing these models. By taking scientific steps, we hope to reduce accidents as much as possible with minimal funding. In order to accomplish this, we must gather and examine a great deal of information about traffic incidents, including the location, time, weather, and characteristics of the roads. Data patterns may be automatically found using machine learning algorithms, which can then be utilized to forecast accident scenarios based on these trends. After that, these models can be used to group mishaps into distinct groups and create safety protocols for each group. This method allows us to create affordable safety precautions that may be used in a range of situations. We think that this strategy might greatly lower the frequency of traffic accidents and increase the safety of pedestrians, drivers, and passengers.

Keywords: Machine Learning, Random Forest, Decision Tree, Logistic Regression, Support Vector Machine

1. Introduction:

To make well-informed judgments about their travel routes, government organizations, corporate sectors, and individual travellers all depend on the availability of accurate traffic flow information [1]. In order to reduce carbon emissions, improve traffic management efficiency, and achieve accurate traffic flow forecast, Intelligent Transportation Systems (ITS) must be implemented [16].

To precisely forecast traffic flow, real-time traffic and historical data gathered from various sensor sources, including inductive loops, radars, cameras, mobile Global Positioning Systems (GPS), crowd sourcing, and social media, are utilized [7]. Due to the massive number of transportation data produced by the widespread use of both new and classic sensors, traffic data has increased dramatically, making transportation management and control increasingly data-driven [12].

Despite the existence of numerous traffic flow prediction models and systems, the majority of them use shallow traffic models and have drawbacks because of the high-dimensionality of the dataset [6]. To satisfy the changing requirements of traffic management systems, public transit systems, and traveler information systems, improvements in traffic flow forecast models and systems are therefore essential [9].

In a variety of fields, including object identification, dimensionality reduction, image classification, and natural language processing, deep learning has become a well-liked method for resolving challenging issues [5]. This is because multi-layer neural networks, which can identify underlying patterns and structures in the data, enable it to learn complex representations of the data [10].

Deep learning is specifically being investigated in the creation of autonomous cars, which have the potential to transform transportation networks by lowering costs and enhancing safety [11]. In order to guarantee safe and effective autonomous driving, researchers and intelligent transportation systems (ITS) are working on driver assistance systems (DAS), autonomous vehicles (AV), and traffic sign recognition (TSR) to give timely and correct information [4]. The use of deep learning in these areas is crucial for enabling the recognition and interpretation of complex visual and auditory cues, and the detection and prediction of objects and events in the environment[3]. Overall, deep learning holds great potential for advancing the capabilities of autonomous systems and enhancing their performance and safety[12].

Due to the vast volume of data involved, traffic flow information prediction is a demanding process that makes it hard to produce reliable forecasts with little complexity [8]. The accuracy of the several algorithms created for this aim is still quite low [2][15]. We suggest using a number of cutting-edge methods, including Deep Learning, Image Processing, Machine Learning, Soft Computing, and Genetic Algorithms, to solve this problem [14]. Numerous research papers and magazines have attested to the effectiveness of these strategies in handling Big Data [13]. By utilizing these strategies, we hope to increase traffic flow prediction accuracy and get past the difficulties posed by big, complicated datasets [17].

2. Purpose Of System

One prevalent issue that can be anticipated with the use of real traffic data is traffic congestion. However, not all users may have easy access to this data, as they frequently need prior knowledge of the most efficient routes to take. In order to solve this problem, real-time traffic prediction using historical and current data sets is required. Numerous causes contribute to traffic congestion, and patterns and trends can be found by comparing different data sets. By using this data, drivers can more efficiently plan their trips by predicting the amount of congestion at various times of the day.

Fuel prices have a big impact on traffic flow as well and can quickly alter patterns of congestion. This forecast's goal is to give real-time traffic and gridlock data, which is crucial for intelligent transportation systems (ITS). However, the intricate traffic patterns found in contemporary cities may be too much for conventional forecast techniques to handle. Consequently, the development of more efficient ITS solutions depends on further study on traffic flow prediction.

3. Block Diagram

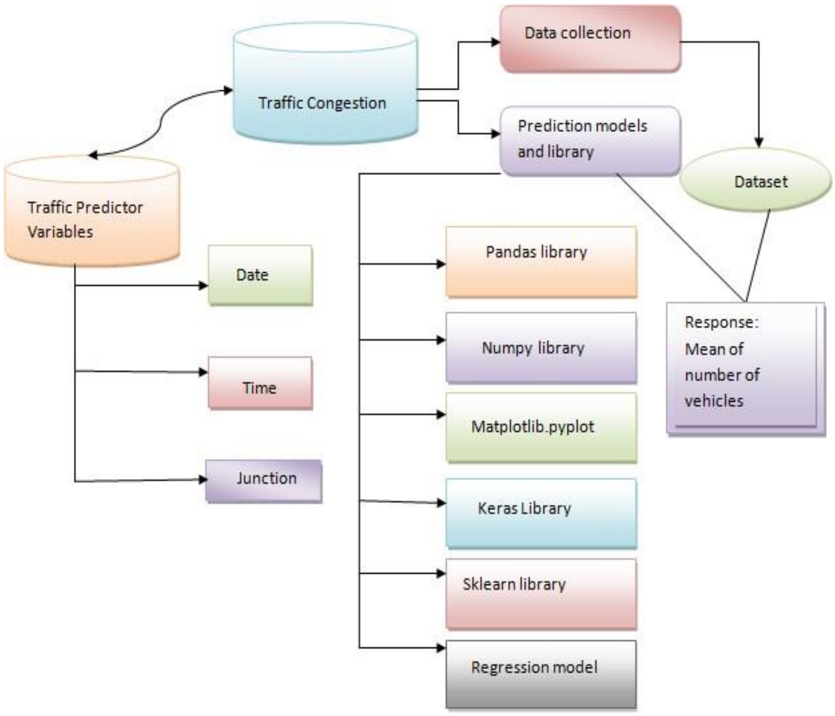

Fig 1 : Traffic Prediction proposed model

4. Methodology

i. We start by taking the dataset.
ii. Filter the dataset in accordance with the specifications and produce a new dataset with the attributes needed for the analysis.
iii. Apply pre-processing to the dataset.
iv. Divide the data into testing and training sets.
v. Use training data to train the model, and then use a classification method to examine testing data.
vi. Lastly, you will receive results in the form of accuracy metrics.

Supervised learning

A model is trained using a dataset with labelled input and output parameters in supervised learning. The labelled dataset serves as the model's training and validation dataset. This method is predicated on the idea of precisely predicting the related input values by using known output values.

A. Classification:

The objective of this supervised learning challenge is to predict discrete values that fall into pre-established classes. The goal is to correctly forecast the class to which an input belongs, and the output has a predetermined selection of labels, such 0 or 1. The capacity of the model to accurately categorize inputs into the appropriate class is used to assess its accuracy. This method can be applied to jobs involving both binary and multi-class categorization. While the model predicts numerous class labels in multi-class classification, it only predicts one class label in binary classification. To classify emails into social, promotional, update, and forum categories, for instance, Gmail employs multi-class categorization.

Example of Supervised Learning Algorithms:

- Gaussian Naive Bayes
- Decision Trees
- Support Vector Machine (SVM)
- Random Forest

Machine Learning | Types of Learning – Supervised Learning

Supervised learning is a machine learning paradigm where the target variable is known beforehand and the algorithm is trained on a labelled dataset. Training a function that can reliably predict the output variable from the input variables is the aim of supervised learning. Supervised learning comes in two main varieties:

Classification: Sorting a new data point into one of the pre-established categories is the goal of classification, a supervised learning technique with a categorical target variable. The system predicts a new instance's class label by learning from a tagged dataset. Image classification, spam detection, sentiment analysis, and medical diagnosis are examples of applications of classification problems. The classification model's performance is usually assessed using measures like F1-score, recall, accuracy, and precision.

Regression is a supervised learning technique that uses input data to forecast a continuous output variable. Finding a function that can predict an output value given an input value is the primary objective of regression analysis. Sales forecasting, weather forecasting, and stock price prediction are a few instances of regression problems. In regression, a labeled dataset including input and output variables is used to train the algorithm. Metrics like mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared) are commonly used to assess the regression model's performance.

A wide range of fields, including speech recognition, computer vision, natural language processing, and medical diagnosis, heavily rely on supervised learning techniques. In order to learn a mapping function that can precisely predict the output variable for a given input variable, these algorithms are trained using labelled datasets. Support vector machines (SVM), random forests, and decision trees are a few of the most widely used supervised learning methods. While random forests employ several decision trees to increase classification accuracy, decision trees create a tree-like model to classify the data. SVM is a linear classifier that uses the best hyper plane to divide the data into distinct classes. These algorithms have produced state-of-the-art outcomes in a variety of applications.

When the data is labelled, supervised learning is a helpful method. But occasionally, the data might not be labeled, or it might be too costly to classify the data. Unsupervised learning, semi-supervised learning, or self-supervised learning may be more appropriate in some situations. Unsupervised learning is the process of learning from an unlabelled dataset with the intention of identifying hidden relationships, structures, or patterns. In contrast, semi-supervised learning enhances the model's performance by integrating both labeled and unlabeled input. Self-supervised learning is a type of unsupervised learning in which the algorithm autonomously generates supervisory signals in order to learn from the data. These techniques have been applied to various domains such as natural language processing, computer vision, and robotics, among others, to address real-world problems where labelled data is limited or unavailable.

SVM

A supervised learning approach called Support Vector Machine (SVM) can be applied to both regression and classification problems. In an N-dimensional space, SVM seeks to identify a hyperplane that can effectively categorize the data points. The number of characteristics in the dataset determines how many dimensions the hyperplane has. The hyperplane is only a line when there are just two input features, and it transforms into a 2-D plane when there are three input characteristics. However, it gets challenging to see the hyperplane when there are more than three features. In such cases, SVM maximizes the margin between the two classes and adds a penalty each time a point crosses the margin. This is known as a soft margin, and the goal of SVM is to minimize (1/margin+∧(∑penalty)) to obtain the optimal hyperplane. Hinge loss is a commonly used penalty that is proportional to the distance of the violation. If there are no violations, there is no hinge loss. SVM has proven to be a powerful tool for classification tasks and has been successfully used in various applications such as image recognition, text classification, and bioinformatics, among others.

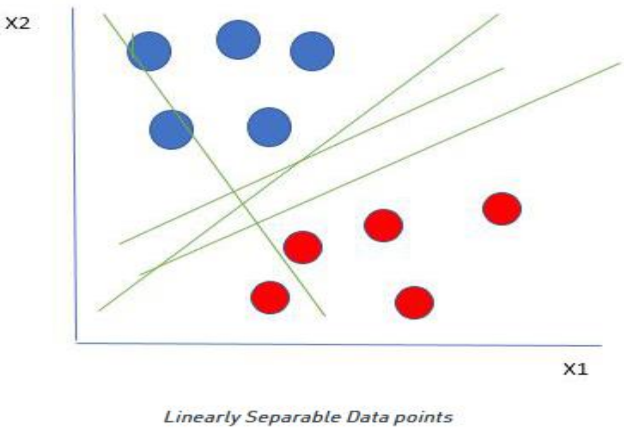

Figure: Datapoints description

4.1 The Data Set

The dataset is the collection of about 1440 Open Data. The following are the dataset attributes

- **i. Day**
- **ii. Date**
- **iii. CodeDay**
- **iv. Zone**
- **v. Weather**
- **vi. Temperature**
- **vii. Traffic**

Day	Date	CodedDay	Zone	Weather	Temperatu	Traffic
Wednesda	1/6/2018	3	2	35	17	2
Wednesda	1/6/2018	3	3	36	16	3
Wednesda	1/6/2018	3	4	27	25	5
Wednesda	1/6/2018	3	5	23	23	3
Wednesda	1/6/2018	3	6	18	42	2
Wednesda	1/6/2018	3	7	11	14	2
Wednesda	1/6/2018	3	8	45	28	4
Wednesda	1/6/2018	3	9	39	18	5
Wednesda	1/6/2018	3	10	25	9	4

Urban traffic congestion has become a major worry due to a number of factors, such as a lack of real-time data, rapid population growth, and poorly coordinated traffic signal timing. The increase in traffic congestion has resulted in significant financial losses, travel time delays, and environmental impacts. To address these issues, traffic flow patterns have been predicted using Python 3 machine learning techniques.

The data utilized in this study came from the Kaggle website, which is a community for people interested in data science. The data consists of two datasets: one from 2015 and one from 2017. Comprehensive information on traffic flow, including time, date, number of vehicles, and intersection details, is included in the statistics. The datasets were collected to facilitate comparison between the two years and evaluate the effectiveness of the machine learning algorithms.

To remove extraneous information and aggregate the remaining data into 1-hour intervals, the collected data was pre-processed. As a result, it was feasible to predict traffic flow precisely each hour. Data transformation, data normalization, and data

cleaning were among the pre-processing techniques utilized to ensure the high caliber and applicability of the data used for training and prediction.
Machine learning techniques used in this study include Random Forest, Gradient Boosting, and XGBoost, among other statistical models. These models were trained on the pre-processed data and then used to predict future patterns of traffic flow. To evaluate the models' accuracy, statistical measures such as mean absolute error (MAE) and root mean square error (RMSE) were employed.
Applying machine learning algorithms may significantly improve traffic flow and reduce urban congestion. With accurate traffic flow pattern prediction, more effective traffic management strategies can be created, such as maximizing traffic signal timing and giving commuters' real-time traffic information. Overall, the findings of this study provide important insights into how machine learning algorithms could be applied to address the issues of traffic congestion in urban areas.

4.2 Regression model

Repressors model analysis could even be a mathematical technique for resolvingthe connection in the middle of one dependent (criterion) variable and one or moreindependent (predictor) variables

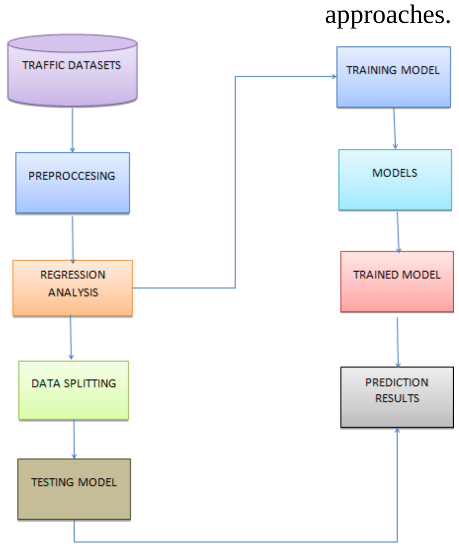

Fig 2: Regression model of traffic prediction

The evaluation yields a fore told value for the benchmark resulting from a sum of scalar vectors of the predictors. The accuracy is measured by computing mean square error. Thus obtaining the expected error from the observed value and also truth value which is equivalent to the standard deviation deployed within the statistical met

5. Results and discussion

A vital aspect of life is the flow of people and products. Travel is expanding rapidly due to the growing population and the need for human social welfare. The number of automobiles is growing daily in tandem with the advancement of technology. Given the speed at which the number of cars is growing, controlling vehicle mobility is essential. Travel time and expense can be optimized with the aid of vehicle management. Accurate background data is necessary for the development of an exact vehicle management system. For the development of an accurate vehicle management system, one of the most crucial pieces of information needed is traffic flow. A survey of current deep learning techniques for traffic flow prediction is presented in this work. Few articles make a significant theoretical contribution, while the majority of contributions are application-based. In order to capture the non-linearity of traffic flow prediction, deep learning models for traffic forecasting have demonstrated encouraging outcomes. Individual traffic flow prediction using deep learning models has a number of benefits, but there are also serious drawbacks. As a result, researchers have recently begun to shift from deep learning architectures to unsupervised and mixed approaches. The numerous deep learning architectures now in use for traffic flow prediction as well as the growing acceptance of hybrid approaches were covered in this review.

To achieve greater accuracy and efficiency, we have evaluated and implemented various machine algorithms. We have employed a decision tree algorithm (DT) to determine categorization and regression.Predicting the value of the target

ISSN: 2456-4265
IJMEC 2025

variables is the aim of this approach. A function that accepts a vector of attribute values as input and returns a single output value called "Decision" is known as decision tree learning.It is classified as a supervised learning algorithm. Both regression and classification problems can be resolved with it. By running a series of tests on the training dataset, DT determines its outcomes.

Another crucial stage for a precise outcome is outlier detection, and to do this, we have employed support vector machines (SVMs), a group of supervised learning techniques that are also applicable to regression and classification. High dimensional spaces benefit from the SVM, which also works when there are less samples than there are dimensions.

One reliable machine learning approach is the Random Forest approach. The term bootstrap aggregation is used to describe it. The random forest technique is mostly used for data classification and is based on predicting models. A single training data set can be used to create many models using the bootstrap approach. Additionally, a bootstrap technique has estimated statistical values using a sample.

The results of performance of the models obtained through different machine learning algorithms that are discussed in this paper. In this table we defined various attributes like Accuracy, Precision, Recall and Time Taken.

Algorithm	Accuracy	Precision	Recall	Time
Decision Tree	88%	88.56%	82%	108.4sec
SVM	88%	87.88%	80%	94.1sec
Random Forest	91%	88.88%	82%	110.1sec

6. Conclusion

A machine learning technique that uses a regression model is what we intend to use to develop a traffic flow forecast system. This technology will forecast traffic flow for the upcoming hour and notify the public of the state of traffic. Users will also be able to find out information about traffic conditions, such how many cars are using a particular intersection. We acknowledge that weather variances, gasoline price fluctuations, and carpooling patterns all impact traffic figures. Therefore, in order to provide precise information on traffic flow, we will compare the projection with traffic data gathered over the previous two years. Users will be able to avoid traffic jams, plan their routes, and make wise decisions thanks to the prediction.

In order to generate a regression model, our traffic flow prediction system will employ a supervised learning algorithm to examine historical traffic data. The traffic flow will subsequently be predicted in real time using this model. Various sources, such as sensors and traffic cameras, will be used to gather traffic data in order to guarantee accuracy. Additionally, the system will consider meteorological conditions that can affect traffic flow, such rain or snow. Furthermore, we will incorporate data on carpooling and fuel prices to offer a thorough examination of traffic trends. The forecast will be presented on an intuitive interface, making it simple for the general public to obtain the data and make well-informed travel decisions. All things considered, our technology will assist in easing traffic congestion, cutting down on travel time, and enhance overall transportation efficiency.

7. Future Work

Our traffic flow prediction system can be improved in the future by utilizing cutting-edge methods like big data, artificial neural networks, and deep learning. In order to give customers precise recommendations for the simplest path to their destination, we will be able to examine additional elements that influence traffic management. There is still room to increase prediction accuracy even if a lot of forecasting techniques have already been used. With more traffic data available, we may create new forecasting models to enhance our forecasts. Our prediction method can assist users in planning ahead and avoiding traffic, which is essential for effective transportation management. In the future, we hope to increase the prediction model's accuracy by creating approaches that are easy to use and accessible, such as integrating weather outlook and GPS data. Additionally, we will highlight accident-prone areas to ensure the safety of our users. We will achieve this through deep learning, big data, and artificial neural networks.

We may also think about integrating real-time data from social media and mobile apps to increase the precision of our traffic flow prediction system. We can use this data to improve our forecasting accuracy and gain greater insights into traffic trends. We can also look into using predictive analytics to foresee shifts in traffic patterns and modify our forecasts appropriately. We will be able to give users more accurate and trustworthy traffic condition information thanks to this.

In addition, we can employ sophisticated visualization methods, such heat maps, to give visitors an understandable depiction of traffic patterns. Users would be able to promptly spot congested locations and adjust their journey as a result.

We can use distributed computing and cloud computing technologies to process massive volumes of data in real-time, ensuring the scalability of our system. This will enable us to manage growing amounts of traffic data and give users precise and timely forecasts.

All things considered, as we integrate new technologies and data sources, our traffic flow prediction system will keep developing and getting better. We can lessen traffic and increase the effectiveness of our transportation systems by giving users fast and reliable information about the state of the roads.

References:

1. Joaquín Abellán, Griselda López, and Juan De OñA. 2013. Analysisoftrafficaccidentseverityusingdecisionrulesviadecisiontrees. Expert Systems with Applications 40, 15 (2013), 6047–6054.

2. Mikhail Belkin and Partha Niyogi. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In NIPS, Vol. 14. 585–591.

3. Ruth Bergel-Hayat, Mohammed Debbarh, Constantinos Antoniou, and George Yannis. 2013. Explaining the road accident risk: Weather effects. Accident Analysis & Prevention 60 (2013), 456–465.

4. Ciro Caliendo, Maurizio Guida, and Alessandra Parisi. 2007. A crash-prediction model for multilane roads. Accident Analysis & Prevention 39, 4 (2007), 657–670.

5. Li-Yen Chang. 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety science 43, 8 (2005), 541–557.

6. Cheng Chen, Manchun Tan. Analysis of traffic flow in incident road section affected by traffic accident . Science Technology and Engineering, 2011; 28.

7. Heydecker B G刿Addison J D. Analysis and modeling of traffic flow under variable speed limits. Transportation Research Part C刿2011;19: 206—217.

8. MccreaJ刿MoutariS. A hybrid macroscopic-based model for traffic flow in road networks. European Journal of Operational Research刿2010; 207:676—684.

9. [US] the United States Traffic Research Committee.Ren Futian, Xiaoming Liu, Jian Rong translation.Road Capacity Manual . Beijing: People's Communications Press, 2007,12.

10. Jiang Liu. Mountain Two-lane road Capacity Research. Beijing: Beijing University of Technology, 2006.

11. Kuanmin Chen, Baojie Yan. Road capacity analysis . Beijing: People's Communications Press, 2003: 44-51,67-91.

12. Jin-chuan Chen, Xiao-ming Liu, Fu-tian Ren, et al.Advances in Operation Analysis of Road Interweaving Area . Highway Traffic Science and Technology.2000,17 (l): 46-50.

13. Jian Rong, Futian Ren, Xiaoming Liu.Study on Simulation Model of Basic Section of Expressway. Beijing: Beijing University of Technology, Ministry of Communications Highway Research Institute. 1999: 1-60.

14. Jinyu Duan, Design of Traffic Simulation Software System for Ramp Area of Freeway. Journal of Highway and Transportation Research and Development, 1999,16 (l): 31-35.

15. People's Republic of China Ministry of Housing and Urban-Rural Construction. Urban Road Engineering Design Code (CJJ37-2012) . Beijing: China Building Industry Press, 2012: 8-9.

16. Yang Weizhong, Zhang Tian. SPSS statistical analysis and industry application case. Beijing: Tsinghua University Press, 2011: 49-51. [13] Structural Equation Modeling.

Chongqing: Chongqing University Press, 2009: 120-150.

17. Qisen Zhang, Yaping Zhang. Road capacity analysis. Beijing: People's Communications Press, 2002: 4-6,36-86 .