IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Traffic Prediction For Intelligent Transport System

J.Nagarshrith , N.Sai Vivan Reddy , M.Samanvay

Ms.K Kusumalatha

Department of Electronics and Computer Engineering,

Sreenidhi Institue of Science and Technology,

Ghatkesar, Telangana, India.

Abstract

As it facilitates traffic management, lessens congestion, and improves the traveller experience, traffic prediction is a vital part of Intelligent Transportation Systems (ITS). In ITS, machine learning (ML) is now a common technology for predicting traffic. In order to estimate traffic for ITS, this paper will explore a variety of ML techniques, including Random Forest, Decision Tree, Logistic Regression, Naive Bayes, KNN, SVM, and Neural Networks. We test these methods using a range of evaluation metrics on real-world traffic datasets, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). We also go through various feature selection methods that are used to increase the precision of traffic prediction. Our findings demonstrate that ensemble methods, such Random Forest and SVM, perform better in traffic prediction for ITS than other methods.

Keywords: Random Forest, Decision Tree, Naive Bayes, KNN, SVM, Logistic Regression.

1 Introduction

One of the key uses of ITS is traffic prediction, which seeks to precisely anticipate the future traffic flow on a certain route. In recent years, the application of ML algorithms to traffic prediction has drawn a lot of attention and shown encouraging results. In order to predict traffic conditions accurately and increase transportation efficiency, machine learning (ML) models can identify trends in traffic data. To increase the efficacy of traffic prediction using ML, a number of issues still need to be resolved. 2 The enormous amount of data needed to train the machine learning models is one of the major difficulties. High-quality, real-time, historical, and external aspects like the weather, events, and road closures must all be included in the data. The reliability of the forecasts is strongly influenced by the data quality. The complexity of traffic patterns, which might vary depending on the time of day, day of the week, season, and location, is another difficulty. Therefore, choosing the right ML model and features is essential for producing reliable predictions. Regression models, neural networks, decision trees, and support vector machines are some of the ML models utilised in traffic prediction. Regression models can be used to model the link between traffic flow and other variables like time of day and weather conditions. Regression models are frequently used for traffic flow prediction. Neural networks can recognise intricate patterns in traffic data and are frequently employed for traffic prediction. Additionally useful for classifying traffic patterns and forecasting traffic flow are decision trees and support vector machines. In conclusion, ITS must include ML-based traffic prediction since it can help solve the issues brought on by traffic congestion. In order to predict traffic conditions accurately and increase transportation efficiency, machine learning (ML) models can identify trends in traffic data. However, there are still a number of issues that need to be resolved, including the lack of high-quality data and the improper choice of ML models and features. Future transport systems may be more effective and sustainable as a result of ongoing research and development in this field.

2 Literature Review

For predicting traffic using intelligent machine learning systems, a number of studies have already been proposed; however, many of these models use shallow traffic models and are still largely failing due to the massive dataset dimension. Here is where our project really shines. Md. Rokebul Islam, Nafis Ibn Shahid, Dewan Tanzim ul Karim, Abdullah Al Mamun, Dr.Md. Khalilur Rhaman, "An Efficient Algorithm for Detecting Traffic Congestion and a Framework for Smart Traffic Control System", Publication on Jan. 31 Feb. 3, 2016 Method: Video processing technique The author of this study proposes a real-time video processing method based on measurements of the volume of traffic on the route. To determine the traffic density, the video clips from a camera are subjected to object detection analysis. To effectively regulate the traffic signal, the calculated vehicle density is compared to that of the other lanes of the route. To assure law enforcement, the system makes use of RFID sensors. Any car that violates the rules of the road can therefore be quickly stopped. In order to provide a development in the current manual traffic control system, this study aimed to do so.[1] G.Monika, N.Kalpana, And Dr.P.Gnanasundari , "An Intelligent Automatic Traffic Light Controller using Embedded Systems ", Publication on, 20th 21st MarchMethod: Wireless Technology They used a density-based traffic control method in this work. It uses an IR sensor as a means of gathering data. These sensors count the number of moving cars and communicate that information to an embedded system. The system uses this data to estimate the times for red and green signals. The automatic traffic management system, which has fixed timing for red and green signals, was also covered in this essay.[2]. Azzedine Boukerche, Jiahao Wang, "A performance 3 modeling and analysis of a novel vehicular traffic flow prediction system using a hybrid machine learning-based model", Published in2020 The author of this paper proposes two novel, inexpensive techniques. One of the methods was created with hardware assistance, whilst the other was created without any hardware assistance. Using computer vision, it determines the number of vehicles on each route, and depending on the number of vehicles, the system determines the optimum running time for green signals and the amount of time for red signals. This system, which is entirely computerised, can switch out the current fixed-time, pre-defined traffic system with one that is dynamically regulated.[3]. Chandana K K, Dr. S. Meenakshi Sundaram, Cyana D'sa, Meghana N Swamy, Navya, "Smart Traffic Management System for Congestion Control and Warnings Using Internet of Things", Publication on May, 2017 This article primarily recommends giving emergency vehicles, such as ambulances, fire trucks, and limousines, priority. Each of these vehicles has an RF transmitter, and the RF receiver is fixed to a signal pole. When there is an emergency, this transmitter sends a signal to the receiver, and the green light turns on for this lane where a pass-ing car is travelling. All lanes, with the exception of the one for emergency vehicles, are lit in red. Data about the current flow of traffic will be gathered. The author of this research used load cells to gather traffic statistics.[4] Azzedine Boukerche, Jiahao Wang,"A performance modeling and analysis of a novel vehicular traffic flow prediction system using a hybrid machine learning-based model", Published in 2020 This research provides an innovative system for predicting vehicular traffic flow that makes use of a hybrid machine learning-based model for precise traffic forecasting. In order to improve traffic flow prediction, the suggested model combines the advantages of two machine learning techniques, namely the Extreme Learning Machine (ELM) and the Long Short-Term Memory (LSTM) network. The system makes use of current and past data, such as weather, events, and road closures. The suggested technique can be utilised to boost overall transportation system efficiency, improve traffic management, and lessen congestion[5] Weibin Zhang,Yinghao Yu,Yong Qi, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning" , Published in 2019 Active traffic control and trip planning are made easier by accurate short-term traffic flow forecasting. The temporal and spatial characteristics of traffic data are not fully utilised by the majority of contemporary traffic flow models. This paper suggests a model for predicting short-term traffic flow based on a deep learning convolution neural network (CNN) architecture. A spatio-temporal feature selection algorithm (STFSA) is used in the proposed framework to choose the best input data time lags and geographical data amounts. Selected spatio-temporal traffic flow features are then retrieved from the real data and transformed into a two-dimensional matrix.[6] "Xiangyang Chen, Ruqing Chen" , "A Review on Traffic Prediction Methods for Intelligent Transportation System in Smart Cities ", Published in 2019 A crucial issue for traffic control and guidance systems, traffic flow prediction is a crucial duty for intelligent transportation systems (ITS). In smart cities, traffic conditions can now be evaluated and anticipated more rapidly and precisely thanks to machine learning and new data sources. Systems for predicting traffic can significantly increase road capacity and ease traffic congestion. In addition to providing a thorough analysis of current traffic prediction techniques for smart cities, this study also examines the 4 issues and difficulties associated with prediction models and identifies potential future developments in short-term traffic flow forecast techniques.[7] .

3 Proposed System

The general process of machine literacy, which involves literacy from data and making prognostications. Then is a high- position overview of the typical way involved 1. Data Collection: The first step is to gather applicable data for your machine learning task. This can involve collecting data from colourful sources, similar as databases, APIs, or manually creating marker datasets. 2. Data pre-processing: Once you have the data, it's important to pre-process and clean it to insure its quality and usability. This step may include handling missing values, removing outliers, homogenizing or spanning features, and garbling categorical variables. 3. point Engineering Feature engineering involves transubstantiating the raw data into a suitable format for machine literacy algorithms. This may include opting applicable features, creating new features through fine operations or sphere knowledge, and transubstantiating the data into numerical representations. 4. unyoking the Data To estimate the performance of the machine literacy model, the data is generally divided into two or three sets training set, confirmation set, and test set. The training set is used to train the model, the confirmation set helps tune hyperparameters and estimate performance, and the test set is used for final evaluation. 5. Model Selection: Next, you need to choose an applicable machine literacy algorithm that suits your task and data. This selection depends on factors similar as the type of problem(bracket, regression etc.), the quantum of data avail- able, and the complexity of the problem. 6. Model Training: In this step, the named model is trained using the training data. The model learns patterns and connections within the data by optimizing its internal parameters grounded on a specified ideal, similar as minimizing the vaticination error. 7. Model Evaluation: After training, the model's performance is assessed using the confirmation set. Common evaluation criteria include delicacy, perfection, recall, F1 score, mean squared error, etc. This step helps identify implicit issues like overfitting or underfitting. 8. Hyperparameter Tuning Machine literacy models frequently have hyperparameters that need to be set previous to training. Hyperparameters control the literacy process and model complexity. ways like grid hunt or arbitrary hunt can be used to find the optimal combination of hyperparameters that maximize performance. 9. vaticination Once the model is trained and estimated, it can be used for making prognostications on new, unseen data. The trained model takes input data and generates prognostications grounded on the learned patterns and connections. 10. Model Deployment: After satisfactory performance is achieved, the trained model can be stationed into a product terrain to make real- time prognostications. This step involves integrating the model into an operation or system that can admit input and give prognostications as affair. It's worth noting that the machine literacy process is frequently iterative, involving multiple cycles of model training, evaluation, and refinement to ameliorate performance and address any issues that arise. The machine learning process involves collecting and pre-processing data, training a model on the data to learn patterns, and using the trained model to make predictions on new, unseen data. It requires steps like fea- ture engineering, model selection, evaluation, and deployment to create an effective and accurate machine learning system. It's worth noting that the machine learning process is often iterative, involving multiple cycles of model training, evaluation, and refinement to improve performance and address any issues that arise.

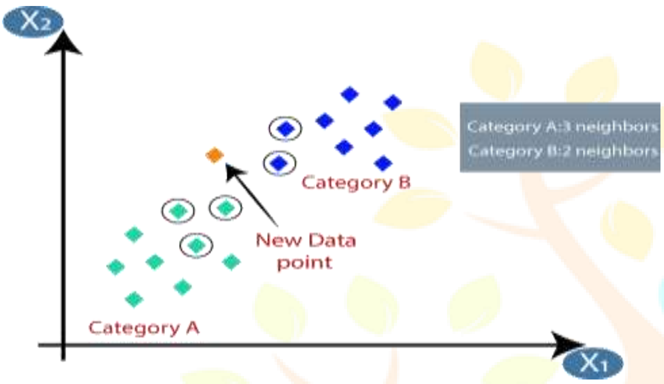
3.1 Naïve Bayes:

A probabilistic classification algorithm called Naive Bayes is employed to estimate the likelihood of an event based on past knowledge. Naive Bayes can be used in the context of traffic prediction to determine, based on past data, the likelihood that a specific traffic occurrence will occur. Naive Bayes can be used, for instance, to forecast the likelihood that traffic would be high at a specific time of day if previous data indicates that peak hours are often when it is heavy. The "Naive" in Naive Bayes refers to the presumption that all features are independent of one another.

$$P(A|B) = P(B|A) * P(A) / P(B)$$

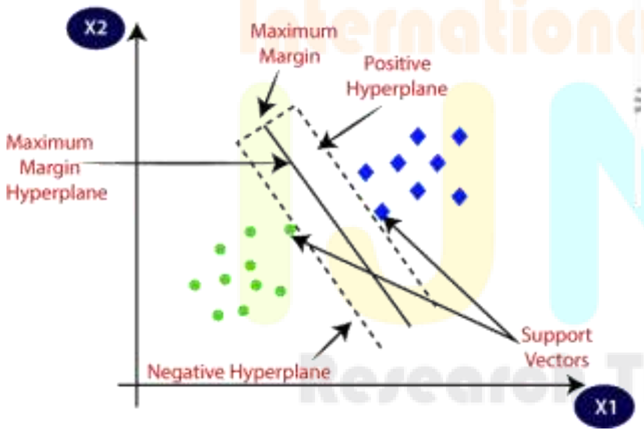
3.2 K-Nearest neighbors(KNN):

K-Nearest Neighbours is a classification technique that sorts data based on the majority class of its neighbours after locating the K nearest data points to the query location. KNN can be used in the context of traffic prediction to forecast traffic conditions based on historical traffic data. KNN can be used to estimate the likelihood of heavy traffic on a comparable day and time, for instance, if data indicates that there was heavy traffic on a specific day at a specific hour.



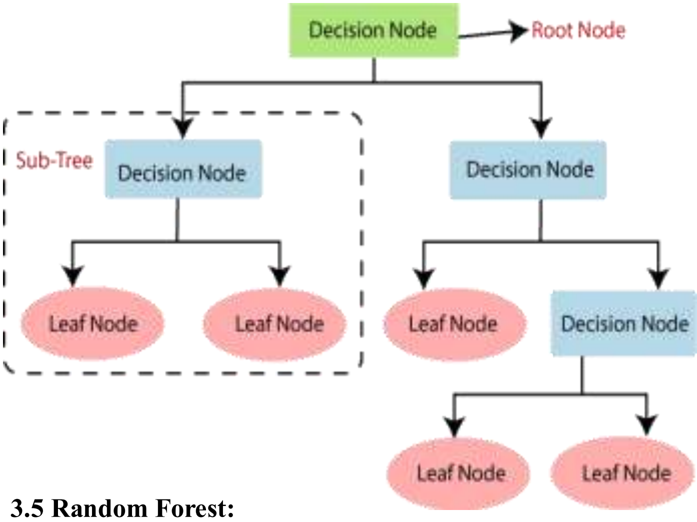
3.3 Support Vector Machine (SVM):

The optimal hyperplane that divides the classes in the data is found by the SVM classification method. SVM can be used to anticipate traffic conditions based on previous data in the context of traffic forecasting. SVM works well with high dimensional data and can handle both linear and non-linear data. Both linear and non-linear data can be handled by the robust SVM method. In order to separate the classes in the data, the optimum hyperplane must be found. In high-dimensional data, SVM works well and can be regularised to prevent overfitting. However, depending on the kernel function used, it may be sensitive and computationally expensive when working with huge datasets.



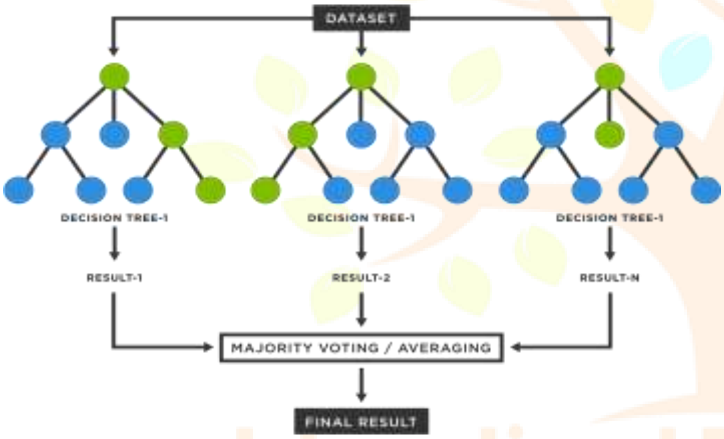
3.4 Decision Tree:

A classification system called Decision Trees builds a tree-like model of decisions and potential outcomes. Decision Trees can be used in the context of traffic prediction to forecast traffic conditions based on historical data. For instance, a decision tree that calculates the possibility of high traffic based on the time of day, day of the week, and weather conditions can be built. Decision trees can be used for both classification and regression applications and are simple to grasp. Both qualitative and numerical data can be handled, and they can also identify intricate non-linear correlations. They can, however, be prone to overfitting, particularly if the tree gets too deep. To solve this problem, ensemble approaches like Random Forest can be used.



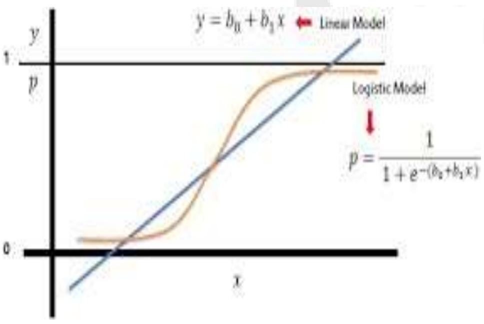
3.5 Random Forest:

The Random Forest ensemble learning algorithm builds numerous decision trees and then combines the output of each tree. Random Forest can be used to predict traffic conditions based on previous data in the context of traffic prediction. High-dimensional data can be handled by Random Forest quite effectively, and it can also manage missing values and noisy data. 6 An ensemble learning system called Random Forest mixes several decision trees to lessen the chance of overfitting. It is capable of handling both category and numerical features and performs well with high- dimensional data. Additionally, it can withstand noisy data and missing values. When dealing with huge datasets, it could be computationally expensive and may not perform effectively when the datasets are unbalanced.



3.6 Logistic Regression:

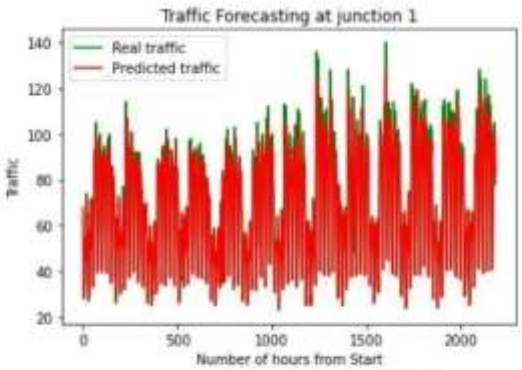
A logistic function is used to model the likelihood of a binary result in the classification technique logistic regression. Based on historical data, Logistic Regression can be used to anticipate traffic conditions in the context of traffic forecasting. The assumption behind logistic regression is that the input features and the output variable are linearly connected. A logistic function is used in the linear process of logistic regression to model the likelihood of a binary outcome. It can handle categorical and numerical data and is easy to deploy. To prevent overfitting, logistic regression can also be regularised. It might, however, struggle with non-linear data and miss intricate correlations between features. All of these machine learning models can, in general, be used to anticipate traffic and can be assessed based on their performance and accuracy. The right model must be selected depending on the available facts and the current challenges.



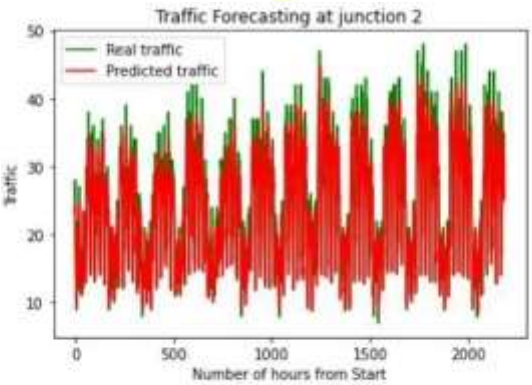
4 Results

The results ofthe prediction made using matplotlib at 4 junctions where green represents real traffic and blue represents predicted traffic.

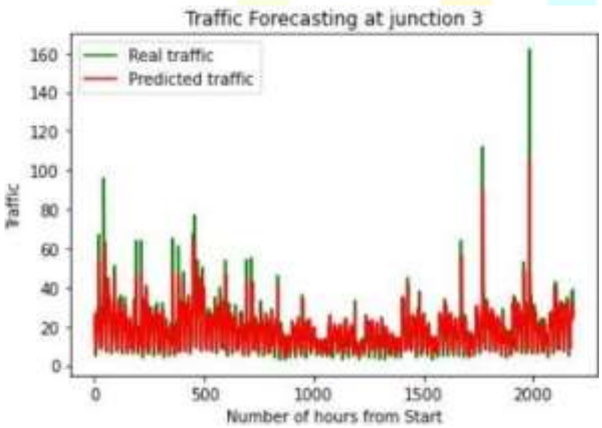
4.1 PREDICTION FORJUNCTION 1



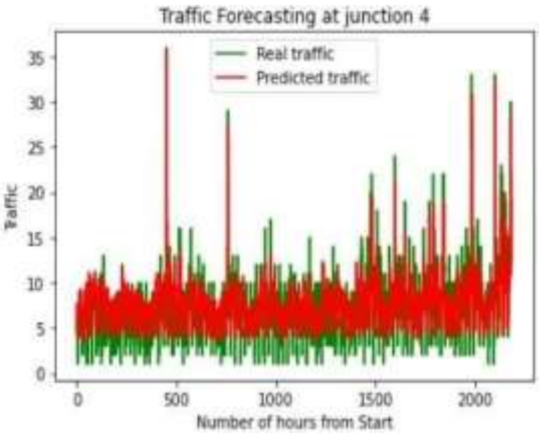
4.2 PREDICTION FORJUNCTION 2



4.3 PREDICTION FORJUNCTION 3



4.4 PREDICTION FORJUNCTION 4



4.5 COMPARISION WITH MODELS

Proposed Models	Accuracy
NAVIE BAYES	75%
KNN	80%
DECSION TREE	87%
RANDOM FOREST	91%

5 Conclusion

It is abundantly evident that machine learning has considerable promise for time series forecasting. This has been demonstrated in both this project and other cited papers. But it's important not to undervalue the effectiveness of current statistical methods. In contrast to ML techniques, the baseline methods did in fact produce respectable outcomes and are quicker to evaluate. ITS is a branch of study and development that integrates quickly developing technology into various platforms for a wide range of cutting-edge applications. The timely collection, processing, and analysis of massive amounts of data become a crucial cornerstone for the successful deployment and run-time operation of many systems. Therefore, developments in ML are regarded as important enabling technologies to fuel an TS revolution. In this study, we looked into how ML is being increasingly suggested as a solution to numerous ITS problems.

6 References

[1] Md. Rokebul Islam, Nafis Ibn Shahid, Dewan Tanzim ul Karim, Abdullah Al Mamun, Dr.Md. KhalilurRhaman, “An Efficient Algorithm for Detecting Traffic Congestion and a Framework for Smart Traffic Control System”, Publication on Jan. 31 ~ Feb. 3, 2016 Method: Video processing technique

[2] G.Monika, N.Kalpana, And Dr.P.Gnanasundari , “An Intelligent Automatic Traffic Light Controller usingEmbedded Systems “, Publication on, 20th& 21st MarchMethod: Wireless Technology

[3] Azzedine Boukerche, Jiahao Wang, ‘‘A performance modeling and analysis of a novel vehicular traffic flow prediction systemusing a hybrid machine learning-based model’’, Published in2020

[4] Chandana K K, Dr. S. Meenakshi Sundaram, Cyana D’sa, Meghana N Swamy, Navya, “Smart Traffic Management System for Congestion Control and Warnings Using Internet of Things”, Publication on May,2017

[5] Azzedine Boukerche, Jiahao Wang,‘‘A performance modeling and analysis of a novel vehicular traffic flow prediction systemusing a hybrid machine learning-based model’’, Published in 2020

[6] Weibin Zhang,Yinghao Yu,Yong Qi, ‘‘Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning’’ , Published in 2019

[7] "Xiangyang Chen, Ruqing Chen" , "A Review on Traffic Prediction Methods for Intelligent Transportation System in Smart Cities ", Published in 2019

[8] https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview

[9]https://www.google.com/search?q=convolutional+neural+network+for+traffic+prediction&rlz=1C1YTUH_enIN1042IN1042&h l=en&source=lnms&tbm

[10] https://hands-on.cloud/wp-content/uploads/2022/01/Implementing-Naive-Bayes-Classification- using-Python1024x562.png

[11] https://static.javatpoint.com/tutorial/machine-learning/images/k-nearest-neighbor-algorithm-for-machinelearning5.png

[12] https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine- algorithm.png

[13] https://www.saedsayad.com/logistic_regression.htm

[14] https://static.javatpoint.com/tutorial/machine-learnng/images/decision-tree-classification- algorithm.png

[15] https://www.tibco.com/sites/tibco/files/media_entity/2021-05/random-forest-diagram.svg

[16] https://www.google.com/search?q=machine+learning+projects&rlz=1C1YTUH_enIN1042IN1042&source=lnms&tbm=isch&sa=X&ved

[17] alfanso "Traffic Flow Prediction for Smart Traffic Lights Using Machine learning published in 2022