IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# TRAFFIC PREDICTION FOR INTELLIGENT TRANSPORTATION SYSTEM

[1]Karangula Navya, [2] Dhanush M, [3]G Chaithanya Reddy, [4]Harshith K, [5]Dhanush M N

[1] Assistant Professor, [2] Student, [3] Student, [4] Student, [5] Student

Department of Information Science and Engineering,

Cambridge Institute of Technology, Bangalore, India

**Abstract:** Machine learning and feature extraction play a very importance in the Internet and health department. The traffic environment consists of everything that can affect traffic on the road, be it traffic lights, accidents, rallies, even road repairs that can cause a large amount of congestion. If we somewhat have imprecise prior information on all of the above and many other everyday situations that can affect traffic, then the driver or rider can make somewhat of an informative decision. Needless to say, it also helps in contributing to the future of autonomous vehicles! In the current decades, traffic data is significantly generated exponentially and we slightly have moved towards embracing big data concepts for transportation. This interesting fact really inspired us to somewhat work on the issue of traffic flow prediction, sort of based on traffic data and models.

Keywords: Traffic prediction, Intelligent Transportation Systems, Random Forest and KNN

## INTRODUCTION`

Various Business sectors and government businesses and person guests require unique and as it should be site visitors float facts. It facilitates the riders and drivers to make higher journey judgment to relieve site visitors congestion, improve site visitors operation performance, and decrease carbon emissions. The development and deployment of Intelligent Transportation System (ITSs) provide higher accuracy for Traffic float prediction. The technique of estimating the amount of vehicle visitors on a road network, generally over a predetermined time frame, is known as traffic prediction. To appropriately predict destiny site visitors styles, it includes inspecting past information on visitors styles, climate, occasions, and other essential variables that affect visitors drift. An Intelligent Transport System (ITS) is a fairly advanced transportation machine that enhances transportation sustainability, performance, and protection via utilizing innovative technology. Traffic control, vehicle-to-automobile and automobile-to-infrastructure communication, wise transportation infrastructure, public transit control, and freight management are only some of the programs and offerings that fall beneath the huge category of clever transportation systems (ITS).The development and deployment of Intelligent Transportation System (ITSs) provide higher accuracy for Traffic go with the flow prediction. It is address as a crucial detail for the fulfillment of advanced visitors control structures, superior public transportation structures, and visitor records structures. The dependency of visitors glide is depending on real-time site visitors and historical records collected from diverse sensor resources, which includes inductive loops, radars, cameras, mobile Global Positioning System, crowd sourcing, social media. Traffic statistics is exploding due to the sizeable use of conventional sensors and new technology, and we've entered the era of a massive extent of statistics transportation. Transportation control and management are now becoming However , there are already a plenty of visitors drift prediction systems and fashions; maximum of them makes use of shallow traffic fashions and are still relatively failing because of the large dataset dimension. Recently, deep getting to know concepts attract many people regarding academicians and industrialist due to their ability to deal with class issues, know-how of natural language, dimensionality reduction, detection of gadgets, motion modelling!!! DL uses multi-layer ideas of neural networks to mining the inherent homes in facts from the bottom stage to the very best degree [4]. integrative advantage of that idea involves time savings. In current decades the lots of attention have made towards the safe automatic driving. It is necessary that the information will be provided in time through driver assistance system (DAS), autonomous vehicles (AV) and Traffic Sign Recognition (TSR). Although already, many algorithms have been developed for predicting the traffic flow.

But these algorithms are not accurate since Traffic Flow involves data having a vast dimension, so it is not very easy to predict accurate traffic flow information with less complexity. The travel time is the essential aspect in ITS and the exact travel time forecasting also is very challenging to the development of ITS. Support Vector Machine (SVM) is one of the most effective classifiers among those which are sort of linear. It is advantageous to prevent overfitting of data!!!

**Literature Review:**

**Panel Azzedine and Boukerche JiahaoWang** explains In recent years, there has been a growing interest in Intelligent Transportation Systems (ITS). Many remarkable applications, such as Vehicular Cloud (VC), intelligent traffic controls, etc., have been proposed under the theme of ITS due to the quick development of vehicular computer hardware, vehicular sensors, and urban infrastructures. These applications have the potential to improve transportation conditions and make them safer, more effective, and more pleasurable. To implement these apps, though, an accurate and effective traffic flow forecast system is required, which presents an opportunity for ITS applications to handle anticipated road conditions. It is also less constrained when it comes to prediction jobs. There are several subclasses within the machine learning technique, such as regression models and kernel-based models. We examine not only the accuracy of various models but also the relevant situation and, occasionally, the particular kind of problem the model was intended to solve. As a result, the goal of this work is to provide a comprehensive and understandable overview of several machine learning models and to evaluate their benefits and drawbacks. This will be accomplished by classifying various ML models according to the ML theory they employ.

**Jiaming Xie** clarifies intelligent transportation systems in clever cities depend heavily on visitors prediction. This article's purpose is to create and execute a site visitors prediction gadget that can as it should be and correctly estimate Hong Kong's traffic float. Finding a way to stability the importance of actual-time and historical visitors records is one undertaking in site visitors prediction. Our ideas integrate records- driven and model-pushed methodologies to leverage both ancient facts and actual-time information. Initially, the constraints of the 2 baseline strategies—the monthly shifting average model and the car-regressive incorporated shifting common—are tested. In order to gain a balance between the two fashions, the hybrid prediction version employs artificial neural networks. The synthetic neural network can weigh between visitors trends diagnosed via past visitors records and traffic records in actual time thanks to neural community education. In addition, a Bayesian community emergency approach is integrated into the prediction method to deal with any emergent situations, including visitors injuries. When predicting the velocity exchange on a link, the emergency prediction approach takes into consideration the site visitors situations of neighboring connections inside the event of an unexpected site visitors state of affairs. Finally, the effectiveness and precision of the advised method are shown by using experimental findings for each short- and long-time period forecasts.

**Yisheng Lv, Yanjie Duan, Wenwen Kang** explains For wise transportation structures to be deployed successfully, visitors glide facts must be timely and correct. We have genuinely entered the generation of massive information for transportation, as seen by means of the explosive growth of site visitors information in current years. Many actual-global programs are still unsatisfied with current site visitors drift forecast strategies, which in most cases use shallow site visitors prediction models. We are inspired via this case to rethink the traffic go with the flow prediction hassle using big site visitors records and deep structure fashions. This studies proposes a completely unique deep- gaining knowledge of-based totally visitors drift prediction gadget that intrinsically takes under consideration the temporal and spatial correlations. Generic visitors float functions are found out using a stacked car encoder version, that's skilled in an insatiable layer-by way of-layer way. To the satisfactory of our understanding, this is the first utility of a deep architecture version for traffic float function prediction that uses automobile encoders as building blocks. Furthermore, assessments show that the suggested technique for predicting visitors go with the flow performs better. The houses of visitors glide beneath extraordinary rainfall eventualities can be learned the usage of the rainfall-included DBN and LSTM.

**Methodology:**

There is more visitors and congestion on the roads due to the boom in vehicles added about by way of the population growth. In order to meet this problem, Transportation Systems (ITS) have created the precise solution that goals to enhance protection, minimize environmental effects, and optimize visitors waft. We are excited to proportion a unique traffic prediction model that accurately estimates site visitors flow by means of utilizing properly-appreciated machine studying methods: Random Forest and K-Nearest Neighbors (KNN).

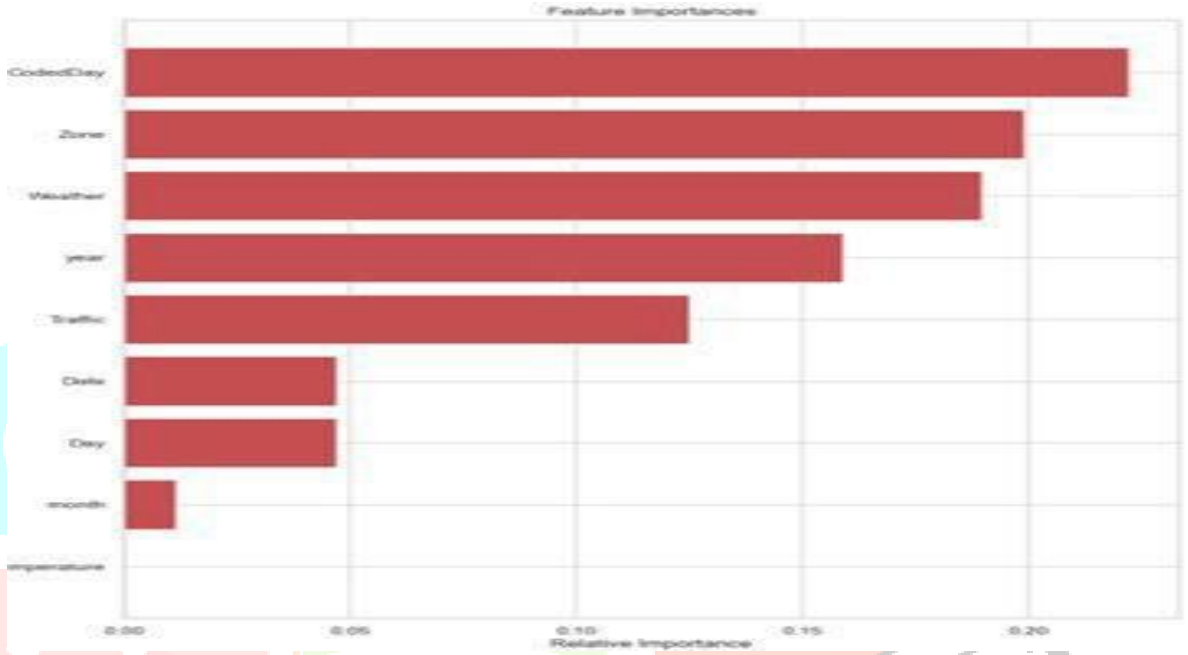


**Fig: Feature diagram**

The traffic dataset is created, the dataset is split, the random forest and KNN algorithms are trained on the data, test data is collected, and accuracy is estimated.

**Dataset Collection**The Kaggle dataset, which includes entities like day, date, zone, is weekend, etc., is where the information needed for the suggested solution was gathered.

**Dataset Splitting**: Around eighty of all the data is utilized for training and twenty percent is used for testing in the proposed method, which employs a simple dataset split. The training set is used to train the Random Forest and KNN algorithms, while the testing set is used to evaluate the model's performance.

**Model Training**: Following the dataset separation process, model training is carried out. The KNN and Random Forest algorithms are fed the training data in this procedure in order to produce a predictive model that can recognize target values on fresh data.

**Model evaluation**: The model is evaluated using various metrics, accuracy, precision, recall, F1-score and confusion matric.

**Testing and validation**: Testing the model with the new data to evaluate the performance of the model and validate that model produce a better accuracy.

**CodedDay**: CodedDay is the most important feature for traffic flow prediction, it is a numerical code assigned to the particular day of the week, this variable is used to represent the day of the week in the dataset. Sunday is assigned the code 0, Monday is assigned the code 1, Tuesday is assigned the code 2 and so on.

**Zone**: Zone is an important feature that refers to the geographical area that is being monitored for traffic patterns.

**Weather :** Weather is an important factor to consider in traffic flow prediction, weather refers to the temperature, rainfall, snow, fog etc.

**Year**: Year is the most important feature for traffic prediction, year is included to provide information about the time period covered by the dataset.

**Date & Day**: Date and Day are not very important features in the traffic prediction but these are also considered while training the model. The date variable provides information about on specific date the traffic data is collected. Day variable provides information about on which day of the week which traffic is collected.

**RESULTS AND DISCUSSION:**

PERFORMANCE TABLE

The below figure shows the train accuracy and test accuracy of the KNN and Random Forest algorithms. This clearly shows that Random Forest model gives the best accuracy when compared to KNN model.

| | Model | Train-Accuracy | Test-Accuracy |
|---|---|---|---|
| 1 | KNN | 1.0 | 0.999467 |
| 2 | Random Forest | 1.0 | 0.999800 |

HYPERPARAMETER

The below graph represents the accuracy vs hyperparameter in which while training and testing for different range of hyperparameter value in range -10 to +10. This graph can be used to determine the range of "C" values (0.0001) that give the model the best accuracy.

```
print('optimal c for which auc is maximum : ',optimal_c)
```


Fig: User Interface for access to predict Traffic

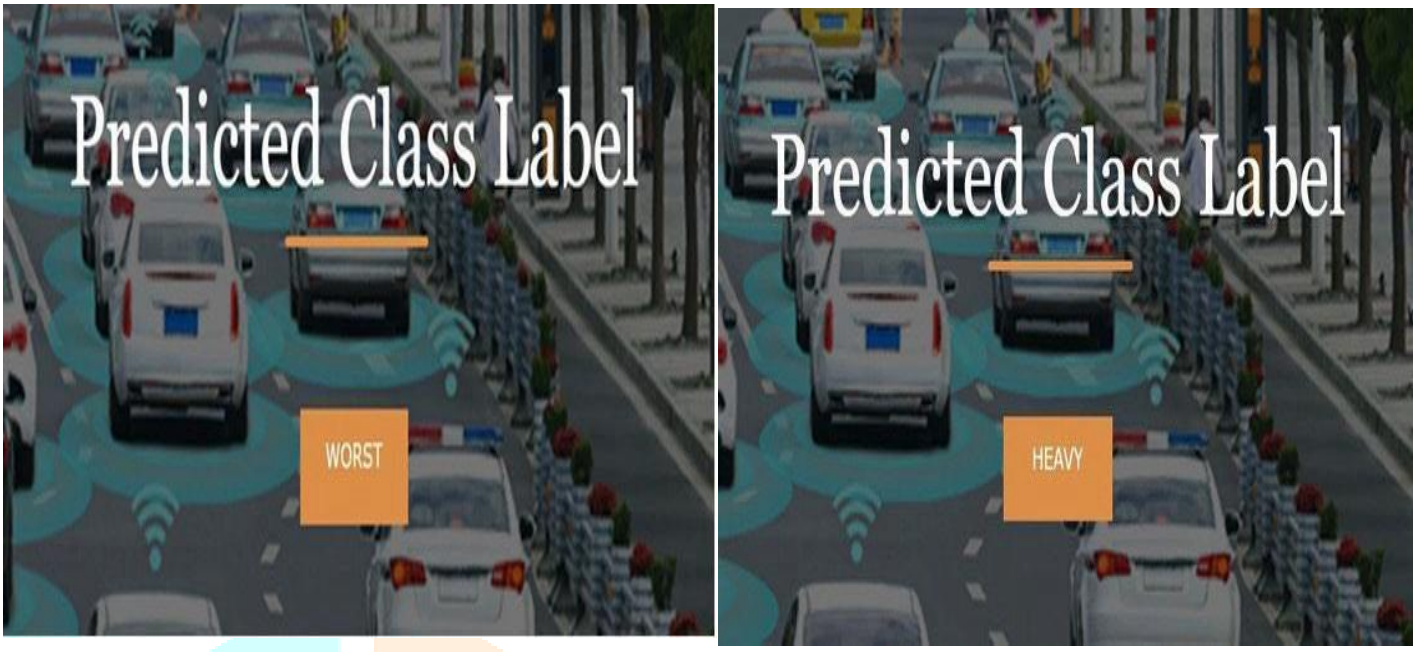


Fig: Analysis-predicted traffic ratio

Giving drivers and transportation planners get right of entry to modern statistics approximately ability traffic congestion will enable them to make informed selections concerning traffic control and course layout. Because this research can predict traffic situations as much as 4 days earlier, it has the ability to reduce tour times and congestion, enhance the safety and efficiency of transportation structures, and assist prevent injuries and different visitors-associated troubles. The purpose is to construct a machine getting to know model which could estimate site visitors conditions for the subsequent four days based totally on prior traffic statistics, weather reports, activities, events, and other applicable elements.

**Conclusion**

In conclusion, research confirmed the effectiveness of the Random Forest and K-Nearest Neighbor (KNN) algorithms for visitors prediction. Random Forest is an extremely good option for site visitors prediction packages because of its adaptability and capability to deal with huge and complex datasets. Conversely, KNN is a easy but powerful set of rules that may provide reliable forecasts with minimal schooling data. The high-quality method will depend on the specifics of the site visitors prediction hassle, due to the fact that each has execs and cons. Taking everything into account, there is a lot of area for development in phrases of the accuracy and performance of visitors prediction models via the use of system studying algorithms together with random woodland algorithm and KNN, which could in the future contribute to stepped forward traffic control.

**References**

[1]. Fei-Yue Wang et al. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. IEEE Transactions on Intelligent Transportation Systems, 2010.

[2]. Yongchang Ma, Mashrur Chowdhury, Mansoureh Jeihani, and Ryan Fries. Accelerated incident detection across transportation networks using vehicle kinetics and support vector machine in cooperation with infrastructure agents. IET intelligent transport systems, 4(4):328–337, 2010.

[3]. Rutger Claes, Tom Holvoet, and Danny Weyns. A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. IEEE Transactions on Intelligent Transportation Systems, 12(2):364–373, 2011.

[4]. Mehul Mahrishi and Sudha Morwal. Index point detection and semantic indexing of videos - a comparative review. Advances in Intelligent Systems and Computing, Springer, 2020.

[5]. Joseph D Crabtree and Nikiforos Stamatiadis. Dedicated short-range communications technology for freeway incident detection: Performance assessment based on traffic simulation data. Transportation Research Record, 2000(1):59–69, 2007.

[6]. H Qi, RL Cheu, and DH Lee. Freeway incident detection using kinematic data from probe vehicles. In 9th World Congress on Intelligent Transport SystemsITS America, ITS Japan, ERTICO (Intelligent Transport Systems and Services-Europe), 2002.

[7]. Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu. Lstm network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, 11(2):68–75, 2017.

[8]. C. Zhang, P. Patras, and H. Haddadi. Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys Tutorials, 21(3):2224–2287, thirdquarter 2019.

[9]. Chun-Hsin Wu, Jan-Ming Ho, and D. T. Lee. Travel-time prediction with support vector regression. IEEE Transactions on Intelligent Transportation Systems, 5(4):276–281, Dec 2004.

[10]. Yan-Yan Song and LU Ying. Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2):130, 2015.

[11]. Bharani B R, “Efficient Technique for Tagging and Archiving the Data Sets and Computation of Performance for Remote Sensing Data using Decision Tree”, Journal of Huazhong University of Science and Technology, ISSN-1671-4512, 2021.

[12]. Preethi S and Maria Kiran L, “Security approach for data migration in a cloud computing environment”, JETIR, (ISSN-2349-5162), 2020.

[13]. Bharani B R, “A survey on predictive analytics and parallel Algorithms for knowledge extraction from date received through various satellites”, IRJET, e-ISSN: 2395-0056, 2019.

[14]. Preethi S and Danya K, “A survey on time and attribute factors combined access control for time-sensitive data in public”, IJEDR, ISSN: 2321-9939, 2017.