TRAFFIC FLOW PREDICTION USING MACHINE LEARNING

CH.Devi[1], V.V.Sai Lavanya[2], T.S.Lavanya Pushpa [3], S.Krishnaveni [4] ,S.Kavya [5] and D.Sai Sri [6]

Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

The effectiveness of the transportation system has recently been improved through the application of numerous traffic prediction systems. An organization called Intelligent Transport System (ITS) works to regulate traffic and creates applications for the future. Traffic flow prediction is useful for controlling traffic flow, traffic lights and travel times. In this application we are taking input from the user ,where we use different machine learning algorithms to predict the traffic flow. This project proposes to compare the various traffic analysis methods like Random Forest Regression, SVM , KNN , GRU and identify the best model by using all methods on a largescale traffic dataset. The top performing model can be applied to future Traffic flow prediction. .

INTRODUCTION

A popular area of study in recent years is Intelligent Transport Systems (ITS). ITS aims to increase transportation security, mobility, productivity, and environmental performance for traffic planners and road users by offering cutting-edge services related to various modes of transport,traffic management and empowering various users to be better informed and make safer, more coordinated, and smarter use of transport networks. Vehicle identification has become more and more important with the ongoing growth of metropolitan roads and the vast construction of motorways. Vehicle detection is a critical ITS task that attempts to provide data for vehicle counting, vehicle speed measurement, traffic accident recognition, and other tasks.

For individual passengers, business sectors, and governmental organizations, accurate and timely traffic flow information is today extremely important. It may aid drivers in making wiser travel choices, lessen traffic congestion, cut down on carbon emissions, and boost traffic operation effectiveness. Providing this kind of traffic flow information is the goal of traffic flow prediction. With the quick development and adoption of intelligent transportation systems, traffic flow prediction has drawn increasing interest (ITSs). It is recognized as a crucial component for the effective deployment of ITS subsystems, particularly advanced traveller information systems, advanced traffic management systems, advanced public transportation systems and advanced commercial vehicle operations. Traffic flow prediction significantly relies on historical and current traffic data gathered from numerous sensor sources, such as inductive loops, radars, and cameras. Due to the widespread use of conventional and innovative sensors, as well as management's increasing reliance on data, traffic data is exploding. There are already numerous systems and models for predicting traffic flow, but because of the large size of the datasets, the majority of them rely on superficial traffic models and continue to have significant limitations.

In general, traffic prediction studies can be categorized into three major categories: naïve methods, parametric methods and non-parametric methods. Naïve methods are usually simple non-model baseline predictors, which can sometimes return good results. Parametric models are based on traffic flow theory and are researched 1separately and in parallel to non-parametric, more data-driven machine learning methods. A strong movement towards non-parametric methods can be observed in the recent years, probably because of increased data availability, the progress of computational power and the development of more sophisticated algorithms. Non-parametric does not mean models are without parameters; but refers to model's parameters, which are flexible and not fixed in advance. The model's structure as well as model parameters are derived from data. One significant advantage of this approach is that less domain knowledge is required in

comparison to parametric methods, but also more data is required to determine a model. This also implies that successful implementation of data-driven models is highly correlated to the quality of available data. In contrast, traffic in urban areas can be much more dynamic and non-linear, mainly because of the presence of many intersections and traffic signs. In such environments, data-driven machine-learning approaches, such as neural Networks, Random Forests and kNN, can be more appropriate, due to their ability to model highly nonlinear relationships and dynamic processes.

LITERATURE SURVEY

Huang Shenghua [1] proposed "Road traffic congestion prediction based on a hybrid random forest and DBSCAN model" .This research suggests a short-term traffic congestion forecast approach based on the random forest algorithm to lessen the inconvenience of people's travel caused by traffic congestion. The level of traffic congestion is determined using DBSCAN, and the historical average speed and traffic flow of urban roads are trained and predicted using the random forest algorithm. This method's accuracy is 94.36%.

Umuhoza Kibogo ,Kong Yan [2] proposed "Evaluating the Performance of LSTM in Traffic Flow Prediction at Different Time Scales". This paper provides specific information on traffic flow. By taking into account various time intervals, it has applied three distinct types of recurrent neural network architecture, including simple RNN, Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU).In this research, an LSTM model for both short and long time intervals is proposed. The accuracy of the predictions has been assessed using two widely used metrics: Mean Absolute Percentage Errors (MAPE) and Root Mean Squared Error (RMSE).

Srihari Nelakuditi [3] proposed "On Selection of Paths for Multipath Routing". This paper tells instead of routing all traffic along a single route, multipath routing schemes divide it up among several paths. It recommends the best path when used in the proportional routing paradigm, where traffic is distributed among a few excellent paths rather than routed. We suggest a hybrid strategy that uses both locally gathered path state metrics and globally traded link state metrics to identify a set of viable paths. Additionally, show that the suggested strategy outperforms other link state update-based schemes in terms of throughput while generating much less overhead.

Zuo Zhang [4] introduced "Using LSTM and GRU neural network methods for traffic flow prediction".This research suggests in the Intelligent Transportation System (ITS), accurate and real-time traffic flow prediction is crucial, particularly for traffic management. In order to forecast short-term traffic flow, we use Long Short Term Memory (LSTM) and Gated Recurrent Units (GRU) neural network (NN) methods. Experiments show that these methods, like LSTM, are based on Recurrent Neural Networks (RNN). GRU is then used to forecast traffic flow.

Dr. H V Kumaraswamy [5] **"Traffic Prediction using Random Forest Machine Learning Algorithms"** This paper uses a variety of machine learning methods to forecast traffic using collected data. With the dataset gathered, the traffic forecast is presented and predicted more accurately than it was with earlier Machine Learning (ML) algorithms. The Random Forest ML algorithms predict traffic statistics with the best accuracy, which is 97.82 percent.Using a variety of sensors, the ITS analyzes car speed and counts the number of vehicles passing it on the road.

PROPOSED SYSTEM

In this project we will be exploring the datasets of four junctions and built a model to predict the traffic on the same. This could potentially help in solving the traffic congestion problem by providing a better understanding of traffic patterns that will further help in building an infrastructure to eliminate the problem. We used different ML algorithms like LSTM, GRU, Random Forest Regressor, KNN to get a more accurate prediction result.

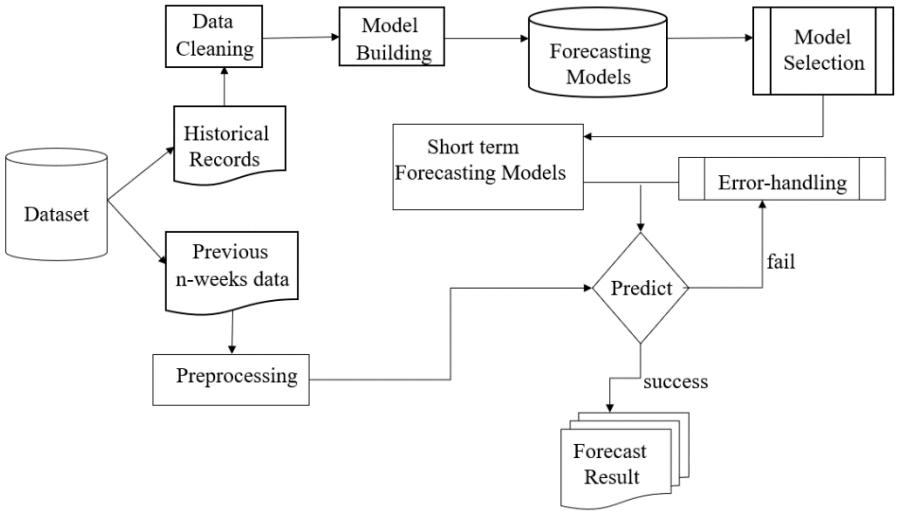


Fig 1: Proposed System Architecture

Machine Learning Models:

a) Random Forest Regression:

An effective supervised learning method is Random Forest, a well-known machine learning algorithm. Both Classification and Regression issues in ML can be solved with it. It is based on the idea of ensemble learning, which is the act of integrating different classifiers to address a complicated issue and enhance the performance of the model. According to what its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the provided dataset and takes the average to increase the predictive accuracy of that dataset." Instead of depending on a single decision tree, the random forest takes the prediction from each tree and guesses the result based on the predictions that have received the most votes.

To forecast the traffic in our project, we employed the Random Forest Algorithm.

1. Choose "k" features at random from the total of "m" features.
   Where k<<m.
2. Using the ideal split point, determine the node "d" among the "k" features.
3. Divide the node into daughter nodes, by using the best split.
4. Continue in steps one through three until "l" nodes have been reached.
5. Create a forest by repeating steps 1 through 4 "n" times to build "n" trees.

```
MAE 19.184727357736243
MSE 602.4551715887698
RMSE 24.54496224459858
RMSLE 3.2005066292973265
R2 Score -0.47824803055755094
#############
```

Fig 2.1: Evaluation Metrics for RFR

b) Support Vector Machine:

Support Vector Machine is a popular Supervised Learning algorithm for Classification and Regression problems. However, it is primarily used in Machine Learning for classification problems. The SVM algorithm's goal is to find the best line or decision boundary for classifying n-dimensional space so that we can easily place new data points in the correct category in the future. A hyperplane is the name given to the best decision boundary. SVM selects extreme points/vectors to aid in the creation of the hyperplane. These extreme cases are referred to as support vectors, and as a result, the algorithm is known as the Support Vector Machine. Multiple lines or decision boundaries can be used to separate classes in n-dimensional

space, but we must find the best decision boundary to help classify the data points. The best boundary is known as the SVM hyperplane. The dimensions of the hyperplane are determined by the number of features in the dataset, which means that if there are only two features (as shown in the image), the hyperplane will be a straight line. And if there are three features, the hyperplane is a two-dimensional plane.We always make a hyperplane with a maximum margin, which means the greatest possible distance between the data points. SVM algorithm can be used for Face detection, image classification, text categorization, etc.

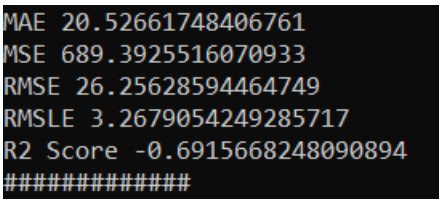


Fig 2.2: Evaluation Metrics for SVM

c) K-Nearest Neighbor:

It is a non-parametric, supervised learning algorithm, which uses proximity to make predictions about the grouping of an individual data point, working off the assumption that similar points can be found near one another. For classification problems, a class label is assigned on the basis of a majority vote i.e. the label that is most frequently represented around a given data point is used. The number of neighbours (k) to consider while making predictions is the most crucial performance deciding aspect of KNN, but there is no specific value of k that works well with all datasets. Lower values of k make the model highly vulnerable to noise and higher values of k make the model very generalized and increase the prediction times. Therefore, the number of neighbours(k)is a hyper-parameter that needs tuning. In order to determine which data points are closest to a given query point, the distance between the query point and the other data points will need to be calculated. This calculation can be performed using various methods but Euclidean, Manhattan and Minkowski distance standout as most used methods. K Nearest Neighbours adds value to our project by learning non-linear decision boundaries which many mathematical models can not capture. It also adapts easily to newly added data, allowing the system to have almost no down time by storing entire training data into memory. Being a lazy learning model, its training time is independent of the number of instances.

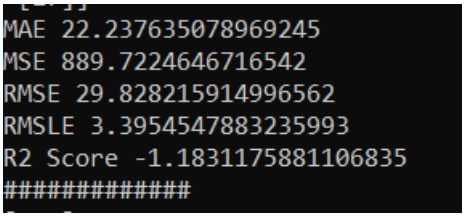


Fig 2.3: Evaluation Metrics for KNN

d) Gated Recurrent Unit:

As compared to lengthy short-term memory, the Gated Recurrent Unit (GRU) is a form of recurrent neural network (RNN) (LSTM). GRU is quicker and requires less memory than LSTM, however LSTM is more accurate when working with datasets that contain longer sequences. Moreover, GRUs handle the issue with vanishing gradients that affects regular recurrent neural networks (values used to update network weights). Grading may become too little to have an impact on learning if it shrinks over time as it back propagates, rendering the neural network untrainable. RNNs can basically "forget" lengthier sequences if a layer in a neural net is unable to learn. The update gate and reset gate are two gates that GRUs utilise to address this issue. These gates can be trained to remember information from further in the past and determine

what information is allowed through to the output. This enables it to transmit pertinent information down a sequence of events in order to provide more accurate forecasts.

```
MAE 19.85447838736492
MSE 648.8988206566916
RMSE 25.473492509993434
RMSLE 3.237638402087733
R2 Score -0.5922071033720588
############
```

Fig 2.4: Evaluation Metrics for GRU

Comparison of the models:

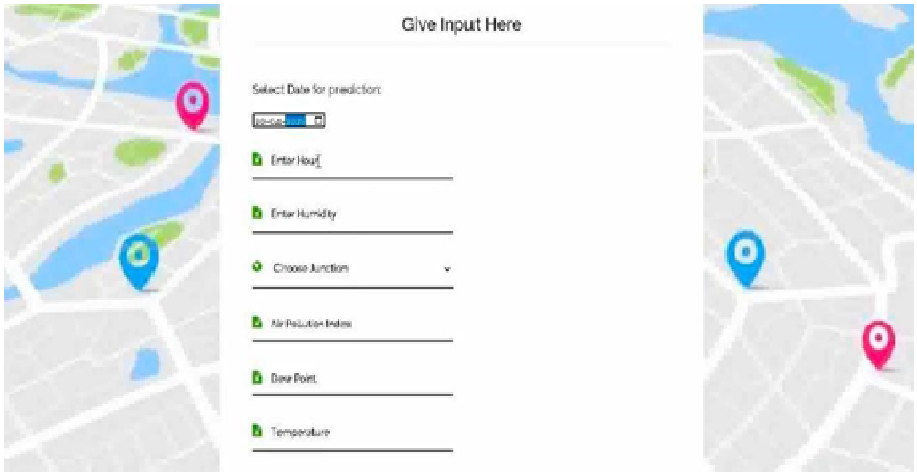
The outcomes of applying several models on the data set are as follows:

MODEL	ACCURACY
Random Forest	88.19%
Support Vector Machine	82.48%
Gated Recurrent Unit	88.04%
k-Nearest Neighbour	83.01%

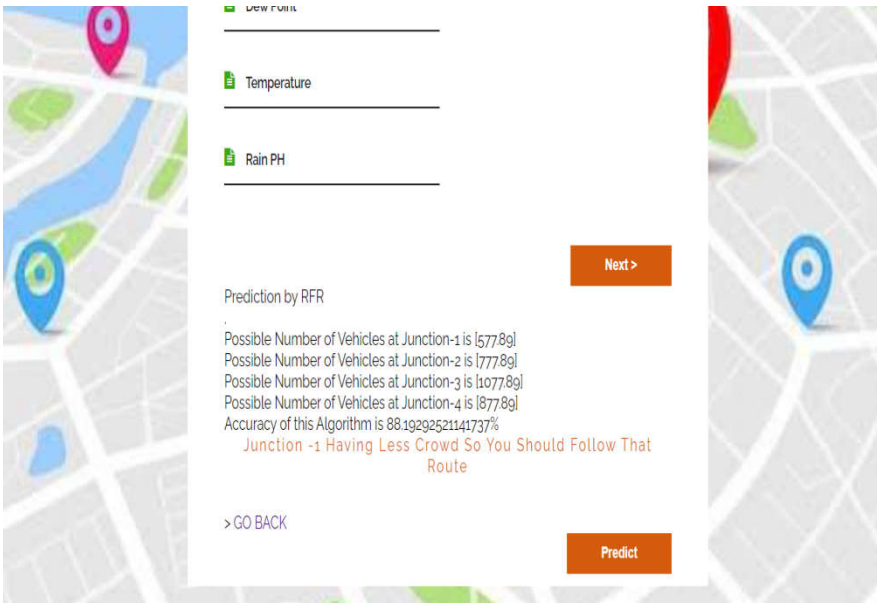
We can infer from the preceding table that the Random Forest technique accurately forecasts the traffic rate by 88.19%.

RESULTS AND DISCUSSIONS

DATA ENTRY



REPORT



CONCLUSION

In this project, a system is developed to analyze and predict traffic using the concept of Random Forest Regression which out performs any of the individual models being used. We used KNN, Gated Recurrent Unit as base models and Support Vector Machine as the meta model to combine the predictions coming from all of the base models. Throughout multiple randomized cross validation we consistently achieved over 88.19% accuracy. Furthermore, we performed a detailed analysis of the traffic and understood the impact caused by the influential features. We documented how various machine learning models are supplementary to each other and how as a random forest can work on fixing the different disadvantages each model has on the given dataset resulting in an overall boost to the system's performance.

FUTURE SCOPE

In Future we will use the sensor data those are installed at various traffic signals and fetching those data and putting inside the datasets and increase the model accuracy dynamically. So that it will be functioned as a complete real time traffic predictor application.

REFERENCES

1. Gaurav Meena, Deepanjali Sharma, and Mehul Mahrishi "Traffic Prediction for Intelligent Transportation System Using ML" February-2020, DOI:10.35940/ijeat.D2426.0410421.
2. Huang Shenghua "Road traffic congestion prediction based on a hybrid random forest and DBSCAN model" May-2020, DOI 10.1109/ICSGEA51094.2020.00075.
3. Umuhoza Kibogo ,Kong Yan "Evaluating the Performance of LSTM in Traffic Flow Prediction at Different Time Scales" Volume 10-March-2021, DOI : 10.17577/IJERTV10IS030115.
4. Srihari Nelakuditi "On Selection of Paths for Multipath Routing" April-2001, DOI:10.1007/3-540-45512-4_13.
5. Zuo Zhang "Using LSTM and GRU neural network methods for traffic flow prediction" volume 10,April-2019,DOI: 10.1109/YAC.2016.7804912.
6. Dr. H V Kumaraswamy "Traffic Prediction using Random Forest Machine Learning Algorithms" volume 11, August-2022,Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com ISSN 2278-6856.
7. Yacong gao ,Chenjing zhou "Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume" August- 2022, Digital Object Identifier 10.1109/ACCESS.2022.3195353.
8. Hainan Wang, Xuetong Wei "Traffic Prediction Using Machine Learning methods" published-

March-2022,DOI: 10.1109/MLBDBI54094.2021.00014.
9. Mouna Jiber, Imad Lamouik ,"Traffic flow prediction using Neural Network" May 2018,DOI:10.1109/ISACV.2018.8354066.
10. Tang C,Sun j, Sun Y,Peng M,Gan N, "A general traffic flow prediction approach based on spatial-temporal graph attention,DOI:10.1109/ACCESS.2020.3018452.