International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 11, Issue 4, July – August 2022 ISSN 2278-6856

Traffic Prediction using Random Forest Machine Learning Algorithms

Ajay C N[1], Dr. H V Kumaraswamy[2]

[1] *Ajay C N Department of Electronics andTelecommunication*
R V College of Engineering Bangalore, India

[2] *Dr. H. V. Kumaraswamy Department of Electronics andTelecommunication*
R V College of Engineering Bangalore, India

Abstract: Numerous traffic prediction algorithms have been used in recent years to increase the effectiveness of the transportation system. Intelligent Transport System (ITS) is an organization that aids in traffic management and develops applications for the future. The ITS analyses vehicle speed and counts the number of vehicles passing it on the road using a variety of sensors. Recent years had the breakthrough in vehicle safety which reduced the traffic accidents. The optimum route during heavy traffic can be determined with the use of traffic prediction, which can save time and gasoline. The different machine learning techniques are employed in this paper that predict traffic based on the data. In comparison to the previous Machine Learning (ML) algorithms using a particular models based on ML the traffic forecast is more effectively displayed and predicted with the dataset collected. The best efficiency of the traffic data prediction is accomplished by the Random Forest ML algorithms is 97.82 percent.

Keywords: ***ITS, SVM, SVR, Random Forest Regression.***

1. INTRODUCTION

In recent years, as the population has increased, so has the number of vehicles on the road. However, greater usage of automobile has brought up a variety of problems, such as higher fuel usage and environmental pollution. For the purpose of resolving the afore mentioned problems, the ITS has supplied particular solutions that have been researched and tested. One of them, concentrates on traffic prediction, which helps to choose the best route prediction while there is traffic.

Vehicle-to-Vehicle (V2V) and Vehicle-to-everything (V2X) communication are two different types of communication techniques used by the ITS institution. Traffic is predicted via V2V and V2X communications among vehicles will give the traffic congestion and help by choosing the best route. The consequence is that the vehicles can more precisely forecast traffic congestion and help drivers choose the best route to a destination, increasing fuel economy and minimizing environmental harm. In the meanwhile, the success of the prediction of traffic using various algorithms by the using of gathered data.

The rapid recent advancements in the computation ability of everyday computers have made it possible to widely apply deep learning methods to the analysis of traffic surveillance videos. Traffic flow prediction, anomaly detection, vehicle re-identification, and vehicle tracking are basic components in traffic analysis [3]. Good solutions to this problem could prevent traffic collisions and help improve road planning by better estimating transit demand.

The study of sophisticated algorithms that may develop automatically over time and with the help of data is known as machine learning (ML). Machine learning algorithms create a model from sample data, also known as training data, to make decisions and predictions without being explicitly prompted to do so in the past. [2] Machine learning algorithms which are used in this paper are Support Vector Machine, Support Vector Regression, Linear Regression, Polynomial Regression, Decision Tree Algorithm, Random Forest Regression.

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which helps in classification as well as Regression problems in Machine Learning. The main aim of the SVM algorithm is to create the best line or decision boundary that can differentiate the n-dimensional space into classes so that the data point will be fitted to exact category. Support Vector Regression (SVR) is a regression algorithm that helps both non-linear and linear regressions. The principle of the Support Vector Machine used in SVR.

Linear Regression is a supervised machine learning algorithm. It tries to find out the best linear relationship that describes the data you have. It assumes that there exists a linear relationship between a dependent variable and independent variable. In linear regression, the algorithm works when the relationship between the real and prediction data is linear. But suppose if we have non-linear data then Linear regression will not able to draw a best-fit line. Hence, the condition will be failed. To solve this problem the introduction of polynomial regression, which helps to draw the curvilinear relationship between independent and dependent variables. The linear model

with different modification will help in increase the accuracy.

The supervised learning algorithms group includes the decision-tree algorithm. It works with output variables that are discrete and continuous. Decision tree regression trains a model in the form of a tree to predict data in the future and generate useful continuous output by observing the properties of an item.

The remaining content of the structure of this article is as follows. In Section II, several learning algorithms are described with references to other papers. The overall design of our proposed work will next be discussed in Section III. The results of the performance review are also shown in Section IV. The conclusion is presented in Section V.

2. DATASET, PRE-PROCESSING AND FEATURE EXTRACTION

The difference between traffic data on the day under study and that on the same day one week before is used as input, performance improves, probably since stationarity in the parameters is better ensured. The tests also indicate a robustness in model performance as the prediction horizon increases, thus suggesting desirable behavior for longer-term prediction

2.1 Dataset

Data collected in this paper are from the GitHub site for the traffic prediction implementations of machine learning methods to show outputs in this prediction system. The data set have the number of vehicles passes on the road with the help of the road sensors. The traffic data collected by the sensors has recorded every hour from 2012 – 2018 also have the climate data with the help of the temperature. The total number of data set which is collected is 48,204.

2.2 Pre-processing and Feature Extraction

Data pre-processing is the most important phase of a machine learning project. Data that shows the important components that are needed to solve an issue is provided in a way that makes machine learning perform best. The table-I shows the pre-processed data after encoding the original data. Data transformation, data reduction, feature selection, and feature scaling techniques used in feature engineering help reorganize raw data into a format suitable for specific kinds of algorithms. By doing this, the computing power and time needed to train a machine learning or artificial intelligence algorithm can be greatly reduced.

From the literature survey for this paper, the observation made for traffic prediction is that many of them used different algorithms like K-Nearest Neighbour (KNN), Recurrent Neural Network/ Artificial Neural Networks (RNN/ANN) and Radial Basis Function (RBF). Among these algorithms the best accuracy achieved by these algorithms for the traffic prediction efficiency is around 94%.

Table-I: Pre-processed data after encoding

	temp	day	date_time	traffic_volume
0	288.28	1	02-10-2012 09:00	5545
1	289.36	1	02-10-2012 10:00	4516
2	289.58	1	02-10-2012 11:00	4767
3	290.13	1	02-10-2012 12:00	5026
4	291.14	1	02-10-2012 01:00	4918
….	….	….	….	….
48199	283.45	2	30-09-2018 19:00	3543
48200	282.76	2	30-09-2018 20:00	2781
48201	282.73	2	30-09-2018 21:00	2159
48202	282.09	2	30-09-2018 22:00	1450
48203	282.12	2	30-09-2018 23:00	954

2.3 Random Forest Regression

Random Forest Regression Algorithm is the next phase of the Decision tree algorithm where the accuracy of the problem will be enhanced with better methods. As, decision tree algorithm only work with a single tree expansion to solve the problem. But Random Forest Regression will create N number of trees and gives the outputs. Further those output will be taken average to get the best accuracy of the problem. The Random Forest Predictor formula is given in equation 1

$$f_{rf}^{N}(x) = \frac{1}{N}\sum_{n=1}^{N} T(x;\ominus_n) \quad (1)$$

Where,
N = total number data used for prediction
x = number of iterations done by the rf algorithm
rf = Random Forest
$\ominus_n$ = characterizes the Nth random forest in terms of splitting into different decision trees.

3. METHODOLGY

The traffic predictor model is designed using the ML algorithms steps as shown in figure 1. Many methods have been considered much time, but in this paper Support Vector Machine, Support Vector Regression, Linear Regression, Polynomial Regression, Decision Tree Algorithm and Random Forest Regression algorithms are used. Among these algorithms the best accuracy of the traffic prediction is found. Initially the traffic data is collected which helps in predicting the traffic by using ML algorithm as defined in problem statement. Now, the data is

pre-processed and the variable and numbers are converted to the numerical values.

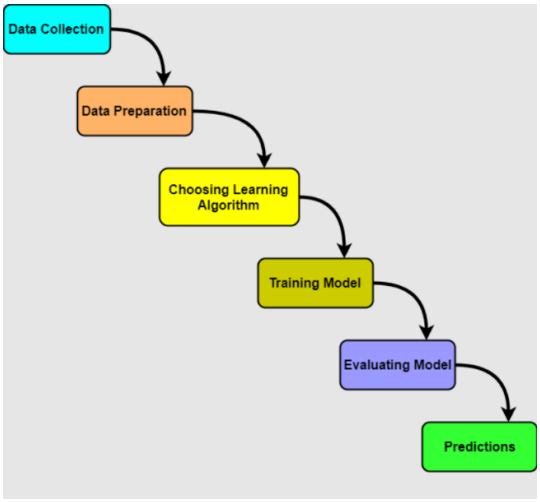

Figure 1 Flowchart of the Machine Learning

The pre-processed data is divided into training data and test data. Next, the ML algorithm is chosen along with training data. The training data will go under the algorithm process and gives us the output. This predicted output will be compared with the test data.

In the final step, the comparison result of the algorithm will give the efficiency of the prediction.

Procedure Run Mechanism

```
#Load Training set
traffic_data <- load(data.csv)
x <- traffic_data[‘Travel Time’’Density’]
 y <- traffic_data[‘Is Road Free’]
#Create Arrays of Training and Test Set
x_train,x_test,y_train,y_test
#Fit Logistic Regression on the Training Set
Running the Regression()
Fit on x_train,y_train
#Predict the Result of the Train Set
prediction<- predict(x-train)
#Compairing the result with test set
Classification_report(y_test,prediction)
```

4. RESULTS

This section discusses about the output of various Machine Learning algorithms using python. The efficiency of the traffic prediction is found by using Mean Square Error (MSE) method where comparison of the test data and the training data output is done for calculating the accuracy of the ML algorithms.

Initially, the Random Forest Regression algorithm has shown 77.4% accuracy to predict the traffic volume as shown in figure 2. After a few trails of the data set with the random forest regression algorithm, the accuracy that achieved was 97.82% to predict the traffic volume as shown in figure 3. To improve the accuracy of the traffic prediction in random forest regression algorithm the groups of multiple decision tree is taken with a particular number for estimators. These exact number will have an exact count on generalized errors which help the algorithms to increase the accuracy of the traffic prediction.

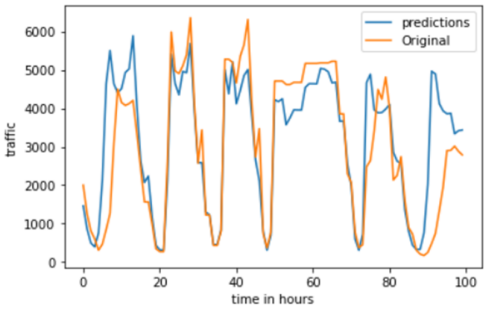

Figure 2 The graph of Random Forest Regression of efficiency of 77.4%

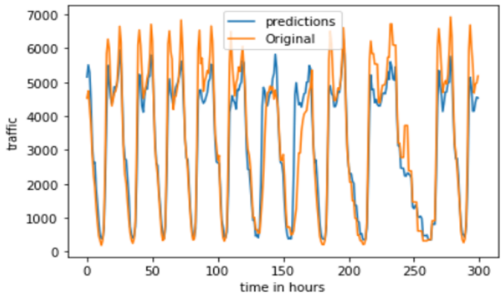

Figure 3 The graph of Random Forest Regression of efficiency of 97.82%

The Table II show the results of accuracy of the traffic Prediction.

Table II: Results of Traffic Prediction.

Algorithms	Efficiency (%)
Linear Regression	80.47
Polynomial Regression	87.47
Decision Tree	85.79
Random Forest Regression	97.82
Support Vector Machine	82.23
Support Vector Regression	90.67

5. CONCLUSION

By observing the different Machine Learning algorithms, the Random Forest Regression algorithm is the best model in predicting the traffic with the given data. The traffic prediction model is implemented 6 different algorithms: Linear Regression, Polynomial Regression, Decision Tree Algorithm, Random Forest Regression, Support Vector Machine and Support Vector Regression. The Random

Forest Regression shows the maximum accuracy as the data used in this process will be more randomized which helps in creating the N number of decision trees with less correlated. Correlating means no two or more-decision tress will have the same output process. Random Forest Regression with hyper parameter tuning with the traffic data has shown the traffic prediction accuracy to 97.82 percent.

ACKNOWLEDGEMENT

This work has been carried out and supported by the Electronics and Telecommunication Engineering Department at RV College of Engineering.

REFERENCES

[1]. G. Xue, Z. Li, H. Zhu and Y. Liu, "Traffic-Known Urban Vehicular Route Prediction Based on Partial Mobility Patterns," 2019 15th International Conference on Parallel and Distributed Systems, 2019, pp. 369-375, doi: 10.1109/ICPADS.2009.129.

[2]. P. Sun and A. Boukerche, "Security Enhancing Method in Vehicular Networks by Exploiting the Accurate Traffic Flow Prediction," 2021 IEEE Wireless Communications and Networking Conference (WCNC), 2021, pp. 1-6, doi: 10.1109/WCNC49053.2021.9417525.

[3]. S. Hua, M. Kapoor and D. C. Anastasiu, "Vehicle Tracking and Speed Estimation from Traffic Videos," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 153-1537, doi: 10.1109/CVPRW.2018.00028.

[4]. C. Manasseh and R. Sengupta, "Predicting driver destination using machine learning techniques," 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2018), 2018, pp. 142-147, doi: 10.1109/ITSC.2018.6728224.

[5]. Yue Dai et al., "Dynamic prediction of drivers' personal routes through machine learning," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850094.

[6]. Iwao Okutani, Yorgos J. Stephanedes, Dynamic prediction of traffic volume through Kalman filtering theory, Transportation Research Part B: Methodological, Volume 18, Issue 1, 2016, Pages 1-11, ISSN 0191-2615.

[7]. Gehrke J.D., Wojtusiak J. (2018) Traffic Prediction for Agent Route Planning. In: Bubak M., van Albada G.D., Dongarra J., Sloot P.M.A. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 5103. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69389-5_77

[8]. Y. Lassoued, J. Monteil, Y. Gu, G. Russo, R. Shorten and M. Mevissen, "A hidden Markov model for route and destination prediction," 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 1-6, doi: 10.1109/ITSC.2017.8317888.

[9]. A. Mariano de Souza and L. Aparecido Villas, "A new Solution based on Inter-Vehicle Communication to Reduce Traffic Jam in Highway Environment," in IEEE Latin America Transactions, vol. 13, no. 3, pp. 721-726, March 2015, doi: 10.1109/TLA.2015.7069097.

[10]. Lo, Chun-Chih & Kuo, Yau-Hwang. (2015). Junction-Based Traffic-Aware Routing scheme for vehicular ad hoc networks. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC. 3001-3005. 10.1109/PIMRC.2015.6666661.

[11]. Lo, Chun-Chih & Kuo, Yau-Hwang. (2015) Traffic-Aware Data Delivery Strategy for Vehicular Ad Hoc Networks Center for Research of E-life DIgital Technology (CREDIT), July 1-3, 2015 ICAST 2015, ISBN 978-89-968650-4-9.

[12]. M. Kim, D. Kotz and S. Kim, "Extracting a Mobility Model from Real User Traces," Proceedings IEEE INFOCOM 2016. 25TH IEEE International Conference on Computer Communications, 2016, pp. 1-13, doi: 10.1109/INFOCOM.2006.173.

[13]. Azzedine Boukerche, Robson E. De Grande, Vehicular cloud computing: Architectures, applications, and mobility, Computer Networks, Volume 135, 2018, Pages 171-189, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2018.01.004.

[14]. P. Sun, N. AlJeri and A. Boukerche, "A Fast Vehicular Traffic Flow Prediction Scheme Based on Fourier and Wavelet Analysis," 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1-6, doi: 10.1109/GLOCOM.2018.8647731.

[15]. Krizhevsky A, Sutskever I, Hinton G E, "Imagenet classification with deep convolutional neural networks," in Proc. NIPS, 2012, pp. 1097-1105.

[16]. Mliki H, Chaari L, Kamoun L, Cousin B, Enhanced ethernet congestion management scheme for multicast traffic. Trans Emerging Tel Tech. 2016;27(11):1563-1579.

[17]. Azzouni A, Pujolle G. A long short-term memory recurrent neural network framework for network traffic matrix prediction.Comput Sci. 2017;3(6):18-27.

[18]. Shaohua, Xu, XueJiwei, and Li Xuegui, "A Sparse Auto Encoder Deep Process Neural Network Model and its System," International Journal of Computational Intelligence Systems, 2017, 10(1): 1116-1131.

[19]. Peiqin Li, JianbinXie, Wei Yan, Zhen Li, GangyaoKuang, "Living Face Verification via Multi-CNNs," International Journal of Computational Intelligence Systems, 2018, 12(1): 183- 189.

[20]. Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Optimized and Meta-Optimized Neural Networks for ShortTerm Traffic Flow Prediction: A Genetic Approach. Transportation Research Part C: Emerging Technologies, Vol. 13, No. 3, 2005, pp. 211–234.

[21]. Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach. Computer-Aided Civil and Infrastructure Engineering, Vol. 23, No. 7, 2008, pp. 536–548.

[22]. Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment AmrElfar, Alireza Talebpour, and Hani S. Mahmassani, National Academy of Sciences: Transportation Research Board 2018.

[23]. Bao G, Zeng Z, Shen Y. Region stability analysis and tracking control of memristive recurrent neural network. Neural Network. 2017;5(1):74-89.

[24]. Rong Fu, Yanmeng Wang and M. S. Berger, "Carrier ethernet network control plane based on the Next Generation Network," 2008 First ITU-T Kaleidoscope Academic Conference - Innovations in NGN: Future Network and Services, 2008, pp. 293-298, doi: 10.1109/KINGN.2008.4542279.

[25]. Ritu, Ritu. (2022). Traffic Management using Machine Learning.

[26]. 26. Ata, Ayesha & Khan, Muhammad & Abbas, Sagheer & Ahmad, Gulzar & Fatima, Areej. (2019). MODELLING SMART ROAD TRAFFIC CONGESTION CONTROL SYSTEM USING MACHINE LEARNING TECHNIQUES. Neural Network World. 2019. 99. 10.14311/NNW.2019.29.008.

[27]. 27. Handanhal Ramachandra, Sujit & Reddy, K. & Vellore, Vivek & Karanth, Sumanth & Kamath, Tareesh. (2016). A novel dynamic traffic management system using on board diagnostics and Zigbee protocol. 1-6. 10.1109/CESYS.2016.7889867.