Traffic Congestion Prediction using Supervised Machine Learning Algorithms

Taiwo, E. O., Ogunsanwo, G. O., Alaba, O. B, Ogunbanwo, A. S
Department of Computer Science, Tai Solarin University of Education, Ijagun, Ogun State
Corresponding author: taemol19@gmail.com

ABSTRACT

Traffic congestion is one of the main severe issues in many cities of the world. Monitoring and understanding traffic congestion is difficult because of its complex nature. The paper analyzed the efficiency of three supervised machine learning algorithms on traffic dataset that was downloaded from Kaggle repository. The dataset consists of 15 attributes and 33750 instances which were further divided into 70% for training and 30% for testing. The dataset was used to formulate predictive models for traffic congestion using three supervised machine learning algorithms: Classification Tree, Support Vector Machine and Ensemble (RUSBoosted) algorithms. The formulation and simulation of the predictive model were carried out using Matrix Laboratory (MATLAB) statistical tool. The results show Classification Accuracy (%) of 99.8, 99.9, 53.9. Prediction Speed (Obs/Sec) of 410000, 10000, 23000 and AUC of 0.66, 0.69, 0.91 for Classification Tree, Support Vector Machine and Ensemble (RUSBoosted) algorithms respectively.

The three models were compared and the best model in terms of accuracy was selected and validated. The study revealed that Support Vector Machine model has higher accuracy, followed by classification tree and Ensemble (RUSBoosted) algorithms. The model is recommended for transport Network services and any other machine learning algorithms can be used for traffic congestion predictive model.

Keywords: Machine Learning, Algorithms, Prediction, Traffic Congestion

INTRODUCTION

Traffic congestion is a crucial challenge since the beginning of road transportation in the busy cities. Adetiloye (2018) defined traffic congestion as a state in transport network wherein the expanded used of road by vehicles in traffic streams generates slower vehicle speeds, time delays as well as a complete paralysis of the traffic network. Traffic congestion is the condition in transport network that is characterized through slower speeds of vehicles (Ramchandra & Rajabhushanam, 2021). Traffic congestion occurs when the numbers of vehicles increase on the road capacity. This leads to loss of valuable time, vehicle wear and tear, high consumption of fuel and accidents. Traffic data are increasing daily, and their analysis is majorly concerned with bringing out predictions of future data. One crucial step in the operation of an intelligent transportation system (ITS) is an accurate traffic prediction system because it will assist traffic managers in the areas of route planning and traffic regulations (Jiber et al., 2020). Monitoring and understanding traffic congestion is difficult because of its complex nature. Traffic congestion is unpredictable. Sometimes it occurs and sometimes it does not, depending on numerous factors.

As a result of present developments in the Artificial Intelligence (AI) discipline and an exponential increase in traffic historical data, AI sub-discipline known as Machine Learning (ML) has a particular set of algorithms that confirmed to be able to predicting traffic data. ML is the study of data and algorithms that a computer system makes use of to enhance its understanding of a given task (Aronsson & Bengtsson, 2019). ML is a branch of Artificial Intelligence that permits computer systems to discover and learn on their own (Jafar Alzubi et al., 2018).ML can also be described as a computational method that trains models to study from experience. ML typically provides systems with the potential to learn and improve from experience without being specially programmed and is commonly called the most famous technologies in the fourth industrial revolution (Sarker et al., 2020; Sarker & Kayes, 2020). Generally, the efficiency of a machine learning solution relies on the nature and characteristics of data and the overall performance of the learning algorithms. Existing literature shows that a wide range of methods have been used for traffic prediction. However, none of the studies get an accuracy of 99.9%. In this study, we forecast traffic congestion and historical traffic data formed the basis for the prediction of future traffic variables. The architecture of the study used to accomplish the traffic congestion predictive model is shown in figure 1.

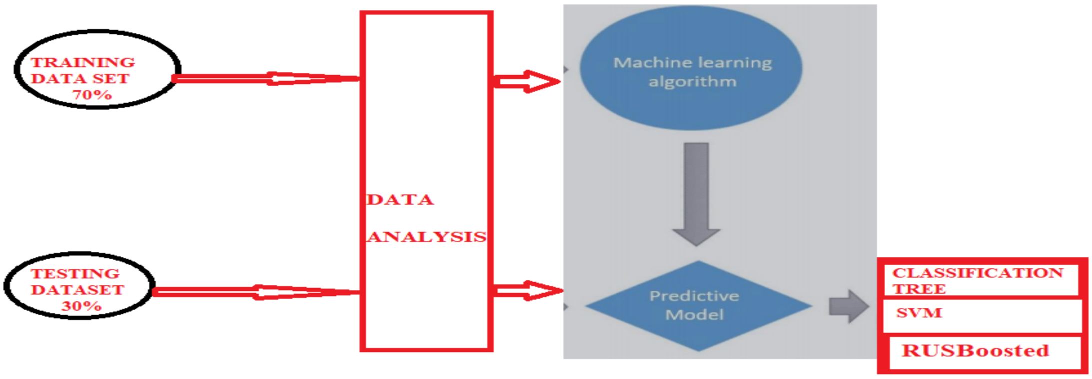

Figure 1 showing the architecture of the predictive model

RELATED WORKS

Majumdar et al (2021) proposed long short-term memory networks for the prediction of congestion propagation across a road network. They employed univariate and multivariate algorithms to build models. Analysis of both univariate and multivariate predictive models showed an accuracy of 84–95%. Liu & Shetty (2021) developed a model for traffic congestion and road accidents. They used four different machine learning algorithms to build models. Experimental results indicated that random forest performed better than the other three algorithms. Rajeev et al (2021) proposed a system for traffic assessment. They utilized machine learning for prediction and regression based algorithm for image detection to analyze the bulk data of the transport system. The system gives greater accuracy in terms of traffic prediction. Geetha et al (2021) developed a system to increase traffic efficiency. They employed genetic algorithm to build model. The system gives much higher efficiency than the existing ones. Qiu & Fan (2021) developed a model to predict travel time on freeway network. They utilized ensemble learning and random forest algorithms to build models. Experimental results showed that random forest produces more accurate travel time in short term prediction than ensemble learning.

Ata et al (2020) developed a TCC-SVM model for an intelligent traffic congestion control. A Support Vector Machine was used to predict congestion with a preprocessing layer to improve the incoming data by dealing with missing values. Evaluation of the results produced from the simulation shows that the performance of the proposed TCC-SVM model is much better when compared to the previous approaches. Zafar and UlHaq (2020) worked on traffic prediction which is based on traffic data collected through Google Map API. Their application is a desktop-based application that predicts traffic congestion state using Estimated Time of Arrival (ETA). They used three machine learning algorithms to build models. The results demonstrate that the random forest classification algorithm has the highest prediction accuracy of 92 percent followed by XGBoost and KNN respectively. Ata et al (2019) proposed a model for smart road traffic congestion. They employed Artificial Back Propagation Neural Networks (MSR2C-ABPNN) to build model. Experiment showed that the model gives attractive results. Kong et al (2019) proposed a prediction network-based on machine learning. They applied long short-term memory (LSTM) on big data-driven traffic flow. They also constructed the LSTM parameter optimization algorithm based on machine learning network. The model has superior performance couple with the increase of training steps. Elfar et al (2018) utilized three machine learning techniques such as logistic regression, random forests, and neural networks for short-term traffic congestion prediction using vehicle trajectories available through connected vehicles technology. Offline and online models were developed in this study. Results indicated that the accuracy of the models built to predict the congested state is 70%.

MATERIALS AND METHODS

Datasets and Attributes: The datasets were obtained from Kaggle dataset repository (https://www.kaggle.com/datasets/bobaaayoung/trafficvolumedatacsv). The dataset consists of 15 attributes namely: date_time, is_holiday, air_pollution_index, humidity, wind_speed, wind_direction, visibility_in_miles, dew_point, temperature, rain_p_h, snow_p_h, clouds all, weather_type, weather_description and traffic_volume.The data consists of information collected from 33750 instances.

Data Analysis: Three different supervised learning algorithms: Classification Tree, Support Vector Machine and Ensemble (RUSBoosted) were predictive model. The values of the 15 features were used to define the input variables of the prediction model. The Matrix Laboratory (MATLAB) statistical tool was used in developing the predictive models. The percentage splits of 70% training and 30% testing were used for the prediction. Each model was compared and the most efficient model was chosen based on evaluation criteria and results.

RESULTS

The results of the traffic congestion model developed for three machine learning algorithms in this paper were shown in Figure 1-3. The evaluation for model were also shown in Table 1-3.

The accuracy of Classification Tree algorithm is 99.8%. It has prediction speed and execution time of 410000 obs/sec and 2.6156 secs respectively. It also has area under curve (AUC) of 0.66 as shown in figure 2.

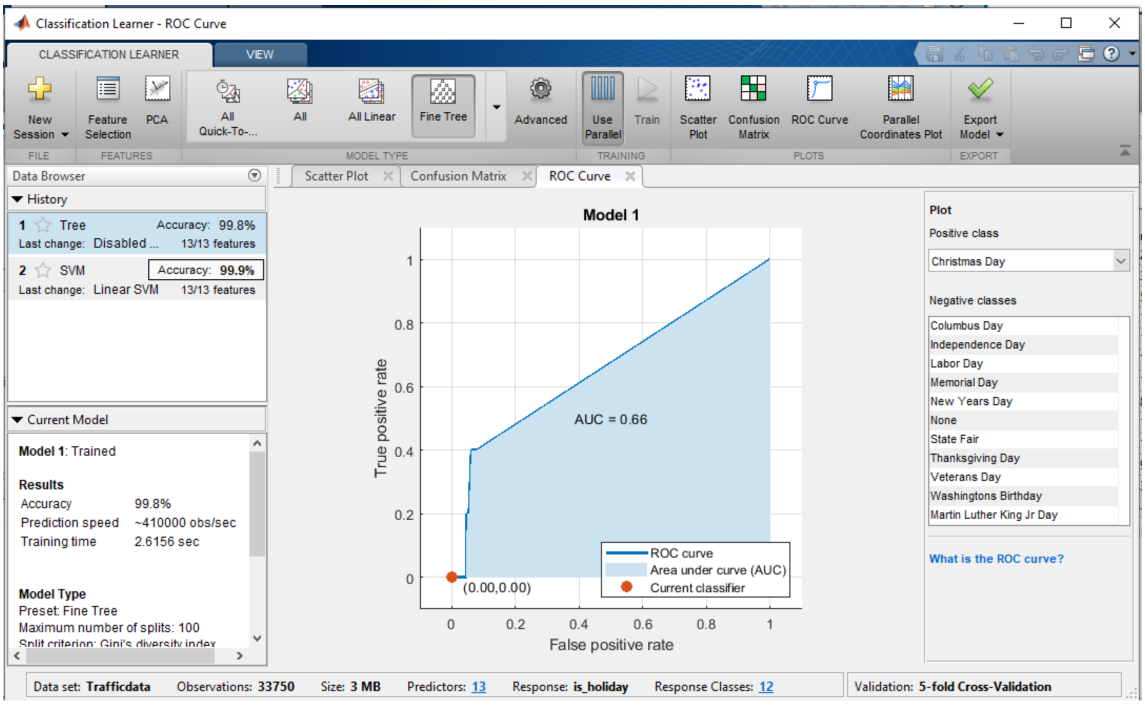

Figure 2: Classification Tree Algorithm for traffic congestion model

Table 1: Results obtained from Classification Tree Algorithm

Number of Instances	33750
Classification Accuracy (%)	99.8
Prediction Speed (Obs/Sec)	410000
Area under curve	0.66
Execution Time (Seconds)	2.6156
Validation	5-fold Cross-Validation

The accuracy of Support Vector Machine algorithm is 99.9%. It has prediction speed and execution time of 10000 obs/sec and 82.493 secs respectively. It also has area under curve (AUC) of 0.69 as shown in figure 3.

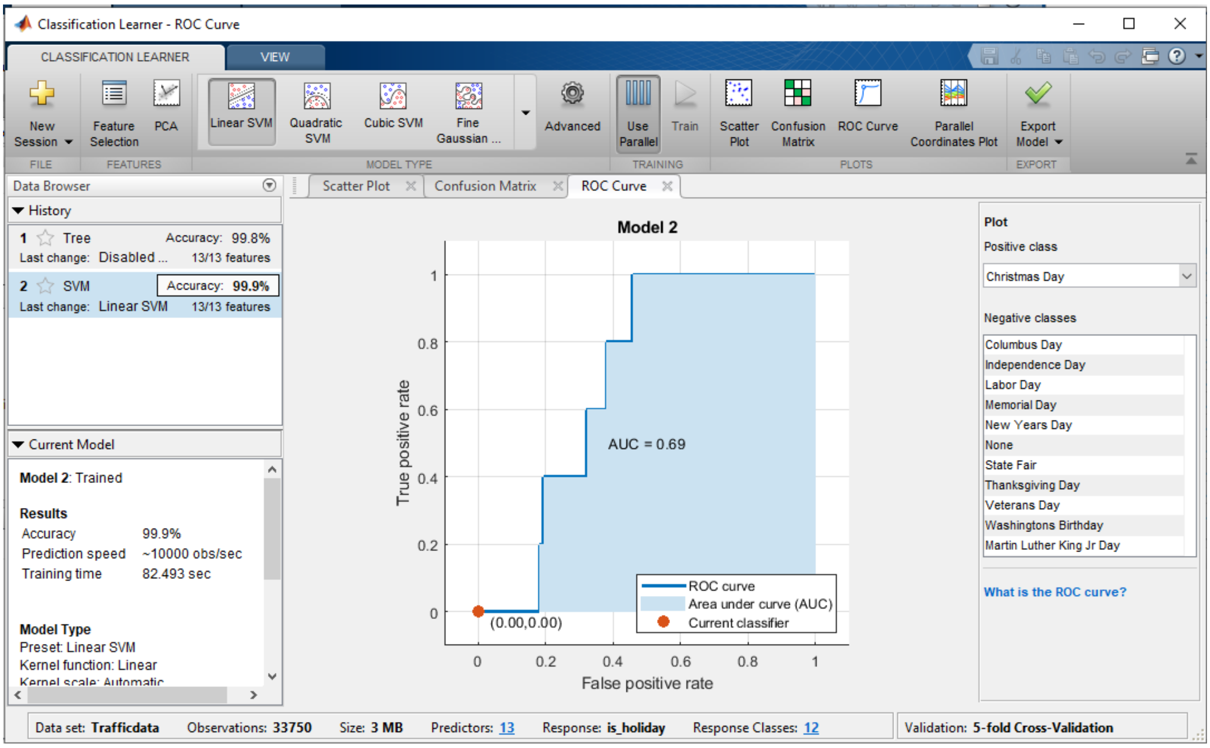

Figure 3: Support Vector Machine Algorithm for traffic congestion Model

Table 2: Results obtained from Support Vector Machine Algorithm

Number of Instances	33750
Classification Accuracy (%)	99.9
Prediction Speed (Obs/Sec)	10000
Area under curve	0.69
Execution Time (Seconds)	82.493
Validation	5-fold Cross-Validation

The accuracy of Ensemble (RUSBoosted) algorithm is 53.9%. It has prediction speed and execution time of 23000 obs/sec and 61.219 secs respectively. It also has area under curve (AUC) of 0.91 as shown in figure 4.

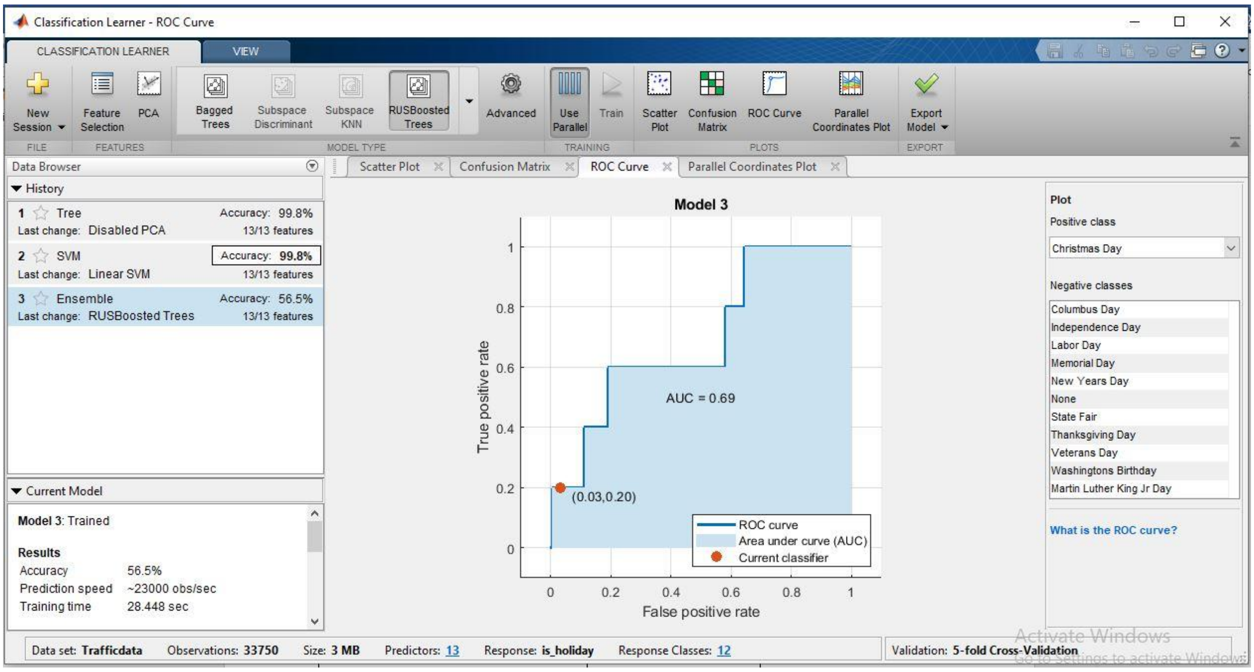

Figure 4: Ensemble (RUSBoosted) Algorithm for traffic congestion Model

Table 3: Results obtained from Ensemble (RUSBoosted) Algorithm

Number of Instances	33750
Classification Accuracy (%)	53.9
Prediction Speed (Obs/Sec)	23000
Area under curve	0.91
Execution Time (Seconds)	61.219
Validation	5-fold Cross-Validation

Table 4: Comparison of results for the three Algorithms used for traffic congestion predictive model

	Classification Tree Algorithm	Support Vector Machine Algorithm	Ensemble (RUSBoosted) Algorithm
Traffic Dataset	**70% 30%**	**70% 30%**	**70% 30%**
Number of Instances	33750	33750	33750
Classification Accuracy (%)	99.8	99.9	53.9
Prediction Speed (Obs/Sec)	410000	10000	23000
Area under curve	0.66	0.69	0.91
Execution Time (Seconds)	26,156	82.493	61.219
Validation	5-fold Cross-Validation	5-fold Cross-Validation	5-fold Cross-Validation

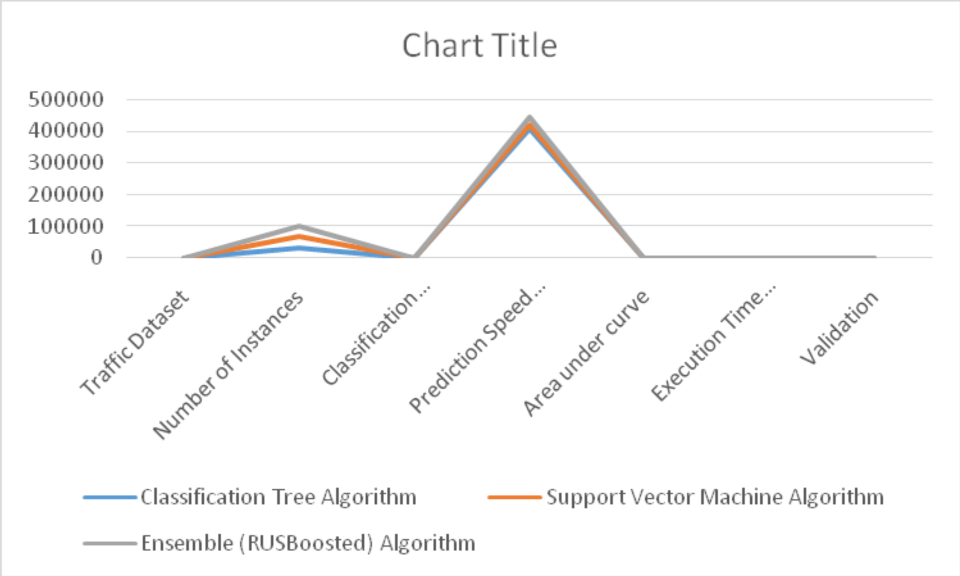

Figure 5: Comparison of the three Machine Learning Algorithms

DISCUSSION

The accuracy of the Support Vector Machine algorithm is higher compared to Classification Tree algorithm and Ensemble (RUSBoosted) algorithm for the traffic dataset divided into 70% 30%. The execution time of the Classification Tree algorithm is faster compared to the other two algorithms as shown in Table 4. With respect to the prediction speed, Support Vector Machine algorithm has fastest speed compare to Classification Tree algorithm and Ensemble (RUSBoosted) algorithm as shown in Table 4. The AUC is 0.66, 0.69, 0.91 for Classification Tree algorithm, Support Vector Machine and Ensemble (RUSBoosted) respectively. Based on the values of the AUC for the three classification algorithms which is closed to 1 shows that the model prediction was right.

CONCLUSION

In this paper, we presented three supervised machine learning algorithms (Classification Tree, Support Vector Machine and Ensemble (RUSBoosted)) to build a model that can be used to predict the traffic congestion. Traffic dataset was split into two, namely: 70% training and 30% testing. The predictive model was implemented on MATLAB statistical tool. Support Vector Machine algorithm recorded the highest prediction accuracy followed by Classification Tree and Ensemble (RUSBoosted) respectively. The accuracy of Classification Tree and Support Vector Machine developed in this paper shows higher predictive rate compared to the accuracy found in the literature Elfar et al (2018) which is 70% accuracy. The model can be used in Transport Network services and other machine learning algorithms can be applied for the Traffic congestion prediction.

REFERENCES

Adetiloye, T. O. (2018). *Predicting Short Term Traffic Congestion on Urban Motorway Networks.* (Doctoral Thesis, Concordia University Montréal, Québec, Canada). Retrieved from https://spectrum.library.concordia.ca/id/eprint/985249/

Aronsson, L. &Bengtsson, A. (2019). *Machine learning applied to traffic forecasting.* A degree project report submitted to Department of Computer Science and Engineering,Chalmers University of Technology and University of Gothenburg, Sweden.

Ata, A., Khan, M. A., Abbas, S., Khan, M. S. & Ahmad, G. (2020). Adaptive IoT empowered smart road traffic congestion control system using supervised machine learning algorithm. *The Computer Journal*, Vol. 00, No. 00, pp. 1-8. https://doi/10.1093/comjnl/bxz129

Ata, A., Khan, M. A., Abbas, S., Ahmad, G. & Fatima, A. (2019). Modelling smart road traffic congestion control system using machine learning techniques. *Neural Network World*, pp. 99-110. https://doi/10.14311/NNW.2019.29.008

Elfar, A., Talebpour, A. &Mahmassani, H. S. (2018). Machine learning approach to short-term traffic congestion prediction in a connected environment. *Journal of the Transportation Research Board(TRR)*, pp. 1-11. https://doi.org/10.1177/0361198118795010

Geetha, V., Gomathy, C. K., Harshitha, T. & Varma, P. V. N. (2021). A traffic prediction for intelligent transportation system using machine learning. *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249-8958 (Online), Volume 10, Issue 4, pp. 166-168. https://DOI:10.35940/ijeat.D2426.0410421

JafarAlzubi, AnandNayyar&Akshi Kumar (2018). Machine Learning from Theory to Algorithms: An Overview. Second National Conference on Computational Intelligence (NCCI 2018). IOP Conf. Series: *Journal of Physics:* Conf. Series1142 (2019) 012012, doi:10.1088/1742-6596/1142/1/012012

Jiber, M., Mbarek, A., Yahyaouy, A., Sabri, M. A. &Boumhidi, J. (2020). Road traffic prediction model using extreme learning machine: The case study of Tangier, Morocco. *Information 2020,* 11, 542; doi:10.3390/info11120542.

Kong, F., Li, J., Jiang, B., Zhang, T. & Song, H. (2019). Big data-driven machine learning enabled traffic flow prediction. Trans Emerging Tel Tech. 2019;30:e3482. https://doi.org/10.1002/ett.3482

Liu, H. & Shetty, R. R. (2021). Analytical Models for Traffic Congestion and Accident Analysis. *Mineta Transportation Institute Publications.* pp. 1-16. https://doi.org/10.31979/mti.2021.2102

Majumdar, S., Subhani, M. M., Roullier, B., Anjum, A. & Zhu, R. (2021). Congestion prediction for smart sustainable cities using IoT and machine learning approaches. *Sustainable Cities and Society* 64 (2021) 102500, pp. 1-10. https://doi.org/10.1016/j.scs.2020.102500

Qiu, B. & Fan, W. (2021). Travel time forecasting on a freeway corridor: a dynamic information fusion model based on the random forests approach. *Smart and Resilient Transport.* Vol. 3 No. 2, 2021 pp. 131-148. e-ISSN: 2632-0495, p-ISSN: 2632-0487. https://DOI:10.1108/SRT-11-2020-0027

Rajeev, G. L., Nancy, R., Megha, S., John, J. M. & John, N. E. (2021). Traffic flow prediction using random forest and bellman ford for best route detection.*International Journal of Engineering Research & Technology (IJERT).* Volume 9, Issue 13, pp. 99-101. ISSN: 2278-0181.

Ramchandra, N. R., &Rajabhushanam, C. (2021). Traffic Prediction System Using Machine Learning Algorithms. Retrieved from https://doi./10.4108/eai.7-6-2021.2308860

Sarker, I. H., Hoque, M. M., Uddin,M. K & Tawfeeq, A. (2020). Mobile data science and intelligent apps:concepts, ai-based modeling and research directions. Mob NetwAppl, pp 1–19.

Sarker, I. H. &Kayes, A. S. M. (2020). Abc-ruleminer: user behavioral rulebased machine learning method for context-aware intelligent services. J NetwComput Appl.page 102762

Zafar, N. &UlHaq, I. (2020). Traffic congestion prediction based on Estimated Time of Arrival. PLoS ONE 15(12): e0238200. https://doi.org/10.1371/journal.pone.0238200