Smart City Traffic Patterns Prediction Using Machine Learning

Chapter · September 2022

DOI: 10.1007/978-3-031-08859-9_10

3 authors, including:

David Oyewola
Federal University, Kashere

Emmanuel Gbenga Dada
University of Maiduguri

Smart City Traffic Patterns Prediction Using Machine Learning

David Opeoluwa Oyewola, Emmanuel Gbenga Dada, and Muhammed Besiru Jibrin

Abstract

Traffic affects every citizen's life in many ways by how long it takes for him or her to travel from home to office, the air condition he or she inhales, the strain generated by traffic jams, sleep, and workouts induced by time spent in traffic. Since motorists cannot see the entire traffic system, the urban traffic system must be anticipated in order to sensitize residents about their mobility choices and the subsequent impact on the environment, as well as to implement smart transport system. The paper used five machine learning models: Bagging (BAG), K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM) to predict traffic pattern in a smart city. The dataset consists of 48,120 rows and 4 columns from which the weekday, year, month, date, and time were extracted. Analysis results show that increase in the number of junctions of the city can alleviate problem being faced on the road by commuters. The Root Mean Square Error (RMSE) of BAG, KNN, MARS, BGLM, GLM are 13.09, 9.23, 23.34, 8.7, and 8.6 respectively. Experimental results demonstrated that GLM attained minimal prediction error compared to other machine learning models such as BAG, KNN, MARS, and BGLM used in this study.

Keywords

Vehicular traffic prediction • Smart cities • Bagging (BAG) • Multivariate adaptive regression spline (MARS) • Bayesian generalized linear model (BGLM) • Generalized linear model (GLM) models

1 Introduction

Among the most difficult problems affecting cities is designing and developing an effective transportation system. As the population of the city grows, so do the city's governmental and non-governmental transport networks, and even minor faults and accidents can have a negative impact on a system that is already stretched to its limits. Traffic jam and delays on transportation systems can cost a city losses in terms economy and have a negative impact on inhabitants' standard of living as they spend a lot of time traveling from one location to another. Traffic gridlock has a number of negative consequences, including lost time, environmental degradation, and safety hazards, as well as a negative influence on economic growth and a deterioration of individuals' relationship with their municipal authorities. It is required that cities address fundamental concerns such as the amount of time drivers spend looking for parking spaces, enhancing highway safety, guaranteeing that public transportation is properly routed, and allowing citizens to travel by many modes of transportation.

The idea of smart cities became a reality thanks to advancement in Information Technology (IT). The use of Advanced Traveler (ATS) and Advanced Traffic Management Systems (ATMS) to effectively managed, control and manage traffic flows is a key element of the smart or intelligent cities of the future. The ATMS/ATS aims to improve the overall performance of the traffic system, for example, to reduce emissions, noise, and travel times. Various types of transport models are routinely employed in the estimation of

D. O. Oyewola · M. B. Jibrin
Department of Mathematics and Computer Science, Federal University Kashere, Kashere, Nigeria

E. G. Dada (✉)
Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria
e-mail: gbengadada@unimaid.edu.ng

D. J. Hemanth (ed.), *Machine Learning Techniques for Smart City Applications: Trends and Solutions*, Advances in Science, Technology & Innovation,
https://doi.org/10.1007/978-3-031-08859-9_10

traffic status. However, the models cannot cover all aspects of the real system and the models must be supplemented with observed traffic status data, for example, traffic numbers and speed/travel measures, in order to have a proper depiction of reality.

Many Internet of Things (IoT) sensors are put across so many sites in a smart city to gather data on traffic, drainage, passenger movement, and so on, and the revelations derived from these data are used to better manage resources, assets, and the like. Many researchers have employed Machine Learning substantially on data collected by IoT sensors in a smart city. The growth of the Internet of Things (Iwendi et al., 2020), as well as big data analytics (Reddy, Reddy, et al., 2020) and machine learning, has made the concept of smart city a possibility.

The goal of a smart city is to deliver good services to its citizens by using modern technology and data analytics on data collected by sensors (Numan et al., 2020). Technology-driven infrastructure, community development programs, smart transportation systems, the use of technology to minimize crimes and burglary (Kumar et al., 2017; Raghavan et al., 2017; Reddy et al., 2019), giving safety to residents, and other factors all contribute to a city's smartness. Machine learning techniques may be used to extract hidden information from data, comprehend patterns in data, and classify or predict data (Gadekallu et al., 2020; Reddy, Rm, et al., 2020; Vinayakumar et al., 2019). Machine learning reduces variability and bias during the model's training and testing phases. As a result, the proposed system solves the limitations associated with the existing techniques used for prediction of traffic pattern. A recursive cloud data service, Artificial Intelligence (AI), and networking means of transportation, populaces, infrastructure, and logistic partners support the public transport system in a smart city. The autonomous waste pickup's sensors monitor all movements in the area of the environment and proactively respond to any possible threats, ensuring citizen safety, and averting accidents before they occur. AI techniques, data analytics, and communication technologies have been effectively used to connect people, roads, and vehicles in order to solve a variety of congested roads problems (An & Wu, 2020).

According to Bhattacharya et al. (2020), astronomical population increase in cities has resulted in an increasing number of individuals having to stay long hours in traffic. For smart city approaches, transportation has become one of the hottest research areas. As a result, the need to develop effective machine learning models that have the capacity to estimate traffic flow in a city. To address this problem, some scholars have proposed different approaches such as machine learning, computer vision, deep learning, and neural networks techniques.

After carefully reviewing some relevant publications, it was discovered that the existing literature does not provide any work that uses BAG, GLM, BGLM, MARS, and KNN techniques to solve the problem of traffic congestion prediction in smart city. Furthermore, the prediction accuracy of some of the existing work is relatively low while in some cases the authors did not use performance metrics to evaluate the performance of their proposed system. Also, some authors did not compare their work with high performing machine learning or deep learning models. Considering these shortcomings, the novelty of this work centers on the use of machine learning for traffic pattern prediction. The summary of our contribution is as follows:

(i) This paper proposed five (5) machine learning techniques for traffic pattern prediction in smart city.
(ii) The Bagging (BAG), K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM) models for solving the problem of traffic congestion prediction in smart city was presented in this paper.
(iii) Minimal Root Mean Square Error (RMSE), Mean Average Square Error (MASE), and Mean Square Error (MASE) for GLM model was attained. The proposed models were evaluated using different performance metrics.
(iv) A recent review of state-of-the-art proposals for traffic patterns prediction in smart city is presented.

The remainder of this paper is organized as follows: The related works in the field of power consumption prediction in smart cities are presented in Sect. 2. The proposed approach used in this study was described in Sect. 3. Section 4 contains the results presentation and analysis, while Sect. 5 contains the paper's conclusion.

2 Related Works

This section presents a brief discussion on recent researches done in traffic prediction in smart cities. It has been explained in section one that ML algorithms have gained wide acceptance among data scientists and researchers for power traffic pattern prediction. This is because of the efficacy of these algorithms in handling prediction problems. These ML models were selected for use in this paper because they are highly non-linear and have capacity to learn from past data. Ozbayoglu et al., (2016) applied regression tree to predict short-term and long-term traffic flow in work zones.

Wu et al. (2007) employed Support Vector Machine (SVM) and a Markov Random Field (MRF) algorithm to design and execute a system that has the capacity to detect parking spaces. The MRF is an implement, a set of vertices with bidirectional edges, that is used to designate a group of random variables with Markov characteristics. Initially, data from parking lot cameras is acquired, and the input images are preprocessed with the isolation of its most important attributes. The SVM is afterward employed to categories parking spaces as either available or unavailable for parking. Finally, MRF resolves any potential SVM classification inconsistencies.

Ng et al. (2019) applied SVM to detect road surface deformities. Yang et al. (2020) utilized the fusion of k-Means and DBN to enhance the traffic network setup. Nallaperuma et al. (2019) employed Deep Neural Networks (DNN) to decide life-threatening road sections that are responsible for volatile traffic situations in cities. The DNN model depicts the relationship patterns of the affected road area and the neighboring road areas that affect movement of vehicles in the specified road section. Verma and Badade (2019) applied machine learning for traffic prediction. The authors used LoRa and a traffic preplanned algorithm. The shortcoming of the work is that they do not present any results or proofs of the effectiveness of their proposed approach.

Williams et al. (2006) proposed five Machine Learning (ML) algorithms for traffic classification predicated on IP (Internet Protocol). The authors demonstrated the usefulness of work accomplished by computer system as an important measure for traffic categorization. Hinsbergen et al., (2007) applied statistical technique for predicting traffic. The drawback of their work is that the proposed system is not designed to solve the problem of traffic congestion.

The authors in Jia et al. (2016) presented a deep learning model that used local data manually fed into the system. Moreover, the traffic prediction done by the system was carried out using deep belief network. Morales et al. (2017) employed machine learning algorithm to examine, cleanse, transform, and model traffic data with the aim of learning valuable information, and making informed decision. The system is then used for traffic predictions. The downside of the proposed work is that the model wastes much energy in inspecting and processing of data.

Zheng et al. (2006) presented a binary neuronal network algorithm for predicting the vehicular traffic. The system predicts traffic flow over a short period of time, to envisage the manner in which the predicted value is likely to influence decision-making or behaviors. The proposed model needs to utilize a Global Positioning System (GPS) for providing information about locations, road traffic conditions, things, predispositions of road users, and so on. The system saves and subsequently transmits the information to users.

3 Methodology

Smart city traffic patterns dataset is collected from the Kaggle (Gosh, 2018) spans from November 2015 to June 2017 as shown in Table 1. In this study traffic information for that period is taken to examine the system performance. The dataset consists of 48, 120 rows with only four columns such as DateTime, Vehicles, Junction, and ID which are insufficient to obtain an accurate result from the dataset. The fundamental issue of limited data is the fact that variation grows with less data (Hoffmann et al., 2019). In order to have sufficient data for out input, we extracted weekday, year, month, day, and time from the DateTime column as shown in Table 2 which displayed the first 10 rows. Our approach involves the use of machine learning techniques such as: Bagging (BAG), K-Nearest Neighbor (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM). The R programming language version 3.6.0 installation and all the necessary packages such as ggplot2, tseries, xgboost, lattice, caret, C50, kernlab, mlbench, randomforest, caretensemble, mass, klar, nnet, mars, and rocr were used for the simulations and generation of the results and graphs presented in the work. The computer system used for the simulations has Window 8.1 64-bits with 6 GB RAM and Intel Pentium dual CPU processor.

3.1 Machine Learning Algorithms

3.1.1 Bagging (BAG)

In bagging model only one prediction may be made from the decisions of different learners. In the case of classification, voting is simply the sum of those decisions. Individual models are created by a variety of methods of bagging and boosting. Models in bagging are given the same weights, whereas more successful models in boosting are given more weighting since an executive may base alternative outcomes on a variety of expert information based on their prior estimations. Individual decision trees are brought together by requiring them to vote on each test. If one class receives more votes than other classes because there are more votes, projections based on a larger number of votes are more reliable (Oyewola et al., 2019). The following are the definitions of bagging algorithms:

Bagging algorithms are defined as follows:

Step 1: Construct a bootstrap sample $(x_1, y_1), \ldots, (x_m, y_m)$ by random drawing m times with replacement from the data $(x_1, y_1), \ldots, (x_m, y_m)$.

Step 2: Compute the bootstrapped estimator $b(.)$ by the plug-in principle: $b(.) = e_m((x_1, y_1), \ldots, (x_m, y_m))(.)$.

Table 1 Description of the dataset

Terms	Meaning	Data
DateTime	A digital record of occurrence of a particular event	Date and time
Vehicles	Different types of transportation plying the road such as car, lorry, or cart	Number of vehicles at different intervals
Junction	A point where two or more things are joined	1–4 junction
ID	ID of each vehicle DateTime	

Table 2 Comprehensive dataset of smart city traffic pattern

Date	Vehicles	Junction	Weekday	Year	Month	Day	Time
01–11–15	15	1	1	2015	11	1	00
01–11–15	13	1	1	2015	11	1	01
01–11–15	10	1	1	2015	11	1	02
01–11–15	7	1	1	2015	11	1	03
01–11–15	9	1	1	2015	11	1	04
01–11–15	6	1	1	2015	11	1	05
01–11–15	9	1	1	2015	11	1	06
01–11–15	8	1	1	2015	11	1	07
01–11–15	11	1	1	2015	11	1	08

Where the function e_m defines the estimator as a function of the data.

Step 3: Repeat steps 1 and 2 N times, where N is often chosen as 50 or 100, yielding $b(.)(k = 1, \ldots., N)$.

The Bagged estimator is

$$b_{bag}(.) = N^{-1} \sum_{k=1}^{N} b^{k}(.) \tag{1}$$

3.1.2 K-Nearest Neighbor (KNN)

A lazy approach is the K-Nearest Neighbor (KNN) classification, who merely keeps trainings, because there is no obvious process of training. It learns via analogy, meaning that a given test tuple is compared with similar training tuples. These tuples should be the closest to the unidentified tuple. A Euclidean distance measures the closest neighbor and the unknown tuple is chosen from its closest neighbors as the most common class. The rate of k can be experimentally determined (Adeniyi et al., 2016). The mathematics equation of KNN is:

$$R(x) = \{\hat{x} | D(x, \hat{x}) \leq d_k\} \tag{2}$$

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{k[y]}{k} \tag{3}$$

$$g(x) = \begin{cases} 1, & k[y=1] \geq k[y=-1], \\ -1, & k[y=-1] \geq k[y=1] \end{cases} \tag{4}$$

where $R(x)$ is the predicting region, closest to the training point x, d_k is the kth order of $D(x, \hat{x})$, $D(x, \hat{x})$ is the distance metric, $k[y]$ is the number of samples in region, labelled as y, $p(y|x)$ is the posterior probability of the observation point x, $g(x)$ is the decision that maximizes the posterior probability.

The basic KNN algorithm is:

1. The value of K used in the KNN algorithm in this work is 3. This enables us to achieve desired accuracy.
2. Calculate the distance of the test sample with all the samples of the training dataset.
3. Sort the distance and define the closest neighbors by minimum K-th distance.
4. Assemble the closest neighborhood categories.
5. Make use of simple majority of the closest neighborhood category as a new data item prediction value.

3.1.3 Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MAR) is a versatile regression method and non-parametric approach that incorporates piecewise linear regression function referred as basic function (bf). In order to estimate performance of

MAR, it uses basic functions (bf) for capturing the hidden non-linear relations between independent input variables (Samadi et al., 2020). Bf is therefore the main component in the generation of a MAR model.

The mathematics equations of MARS:

$$y = c_o + \sum_{m=1}^{M} c_m bf_m(x) \tag{5}$$

$$bf_m(x) = \max(0, x - t) \tag{6}$$

Or

$$bf_m(x) = \max(0, t - x) \tag{7}$$

where y is the output of the MARS model, c_o is a constant value, c_m is the coefficient, x is the input variable, t is the threshold, and bf_m is the basis function.

3.1.4 Bayesian Generalized Linear Model (BGLM)

The Bayes Generalized Linear Model (BGL) assumes preliminary or prior distribution based on preliminary or prior information and subsequent distribution is achieved through the integration of sample information with such prior information. In general, information collected from the post or posterior distribution is closer to true information, since it brings together sample data and expert views (Shi et al., 2019).

The prior distribution is:

$$\pi(\beta, \theta) \propto \exp\left(-\sum_{i=1}^{n} n_i \theta_i^{-1} - \sum_{j=1}^{p} \frac{\left(\beta_j - u_{\beta_i}\right)^2}{2\sigma_{\beta_j}^2}\right) \Pi_{i=1}^{n} \frac{\theta_1^{-1-m_i}}{\sqrt{\sigma_{\beta_j}^2}} \tag{8}$$

The posterior distribution is:

$$\pi(\beta|\theta, Y) \propto L(Y|\beta, \theta)\pi(\beta, \theta) \propto \Pi_{i=1}^{n} \frac{\theta_i^{\theta_i exp \sum_{j=1}^{P} x_{ij}\theta_j)}}{\tau\left(\theta_i \exp\left(\sum_{j=1}^{P} x_{ij}\beta_j\right)\right)}$$
$$y_i^{\theta_i \exp\left(\sum_{j=1}^{P} x_{ij}\beta_j\right)^{-1}} \exp(-\theta_i y_i) \exp\left(-\sum_{i=1}^{n} n_i \theta_i^{-1} - \sum_{j=1}^{P} \frac{\left(\beta_j - u_{\beta_j}\right)^2}{2\sigma_{\beta_j}^2}\right) \Pi_{i=1}^{n} \frac{\theta_i^{-1-m_i}}{\sqrt{\sigma_{\beta_j}^2}} \tag{9}$$

where β, θ are independent and $\beta_j (j = 1, 2, \ldots, p)$ follows normal distribution $N\left(u_{\beta_j}, \sigma_{\beta_j}^2\right)$.

3.1.5 Generalized Linear Model (GLM)

Generalized Linear Model (GLM) is a package that adapts linear and similar models with a penalized maximum probability. The regularization path is calculated on a grid of values for the lambda regulation parameter for the lasso or elastic net penalty (Friedman et al., 2010). The approach is very quick and can make the input matrix × sparse. It adapts to model regressions of linear, logistic and transnational, fish and Cox. It can also fit linear regression with multiple responses, linear generalized patterns for individual families and lasso models. GLMNet was developed by Jerome Friedman, Trevor Hastie, Rob Tibshirani, Balasubramanian Narasimhan, Kenneth Tay and Noah Simon and Junyang Qian (Simon et al., 2011).

The mathematical equation of GLM is:

$$f(y_i|\theta_i) = \exp\{\theta_i y_i - \vartheta(\theta_i) + c(y_i)\} \tag{10}$$

$$\rho_i = x_i^{'}\beta \tag{11}$$

$$\theta_i = g(\rho_i) \tag{12}$$

where $x_i^{'} = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$ is $1 \times p$ vector matrix of covariates X, $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)^{'}$ is a p vector of regression coefficient, $\vartheta(.), c(.)$ are known function and $g(.)$ is a monotonic differentiation function.

3.2 Performance Evaluation

The accuracy test of the smart city dataset is evaluated using:

A. *Mean Absolute Error (MAE)*

Consider a set of the actual closing price C_p and the predicted values $\widehat{C_p}$. MAE are given as follows:

$$\frac{1}{n}\sum_{n=1}^{n}\left|C_p - \widehat{C_p}\right| \tag{13}$$

B. *Root Mean Square Error (RMSE)*

RMSE is given as:

$$\sqrt{\frac{1}{n}\sum_{n=1}^{n}\left(C_p - \widehat{C_p}\right)^2} \tag{14}$$

C. *Mean Square Error (MSE)*

MSE is given as

$$\frac{1}{n}\sum_{n=1}^{n}\left(C_p - \widehat{C_p}\right)^2 \tag{15}$$

D. *Mean Absolute Scaled Error (MASE)*

MASE is given as

$$\frac{1}{n}\sum_{n=1}^{n}\frac{\left|C_p-\widehat{C_p}\right|}{\frac{1}{n-m}\sum_{n=m+1}^{n}\left|C_p-\widehat{C_p}\right|} \tag{16}$$

where m is the seasonal period of the closing price and n is the trading days.

3.3 Proposed Traffic Pattern Prediction System

The parameters settings which consist of values used for the different variables in the different models in this paper are presented in Table 3.

Figure 1 depicts the block diagram of the proposed traffic patter prediction in a smart city.

The proposed Smart city traffic pattern was borne out of the problem associated with urban city. Smart cities use big data for decision-making and a revolution in the Internet of a Thing (IoT) allows such data to be collection and transmission of such data. Transportation is a fundamental element influencing metropolitan areas and a crucial use case for smart cities. Smart transportation technologies, such as GPS Bus, traffic cameras, taxis, and geospatial technology, which is frequently employed in a traffic control center to monitor and coordinate a wide network of sensors, are discovering creative methods to reduce traffic congestion and urban mobility. Smart city traffic pattern incorporates all the big data obtained from GPS Bus, traffic camera, GPS Taxi, and geospatial technology. The traffic pattern system consists of data visualization, data cleaning to remove unwanted and other irrelevant values. However, the predicting modeling is the ware house of machine learning such as KNN, BAG, MARS, BGLM, and GLM used in this research. The accuracy of the machine learning is evaluated in the performance evaluation section.

4 Results and Discussion

In order to accomplish sustained urban growth, Smart City is described as a multifaceted strategy integrating multisector and multinational players, structured into new technologies to address social, economic, and environmental issues (Ruhlandt, 2018). The conditions of road traffic have to be collected in order to manage and control traffic flows. The traffic condition can be characterized on a given section of the road with speed, flow, and density. In this study, traffic management is of paramount importance in a smart city. Computational experiments were conducted with the data obtained from Kaggle.

Figure 2 shows the exponential increases in the number of vehicles between 2015 and 2017. The number of vehicles plying the road is at its peak in 2016. This shows that there will be increase in the traffic congestion during this period.

Figure 3 displays the increase in the number of junctions between 2015 and 2016. There is exponentially increase in the number of junctions every year. By comparing Figs. 2 and 3, we observed that as the number of vehicles increases, government also increases the number of junctions which leads to the reduction of traffic between 2016 and 2017. Congestion in traffic is characterized in transportation by slower speeds, longer travel durations, and an increase in vehicle queuing.

Figure 4 shows the longer trip times experience on the road and vehicular queues from 1 November, 2015 to 31 December 2015. This shows that the vehicular queuing increases with time.

In 2016–2017, the number of vehicles increases to 54,000 against 5700 obtained in 2015 which leads to increase in vehicular movement in (Morning, Afternoon, Evening, and Noon) time but as the number of junctions increases in the city, it resulted into decrease in vehicular queuing in 2017 to 48,000 as also shown in Figs. 5 and 6. The study reveals that increase in the number of junctions on a high way may lead to reduction in vehicular movement on the road.

Table 3 Parameter setting for BAG, KNN, MARS, BGLM, and GLM models

MODEL	RMSE
BAG	nbag = 100, comb = NULL, coob = TRUE
KNN	K = 3
MARS	Penalty = 2, degree = 1,pmethod = “forward”, nk = 200
BGLM	Family = Gaussian, prior.mean = 0, prior.scale = NULL, prior.df = 1, maxit = 50
GLM	Family = binomial, optim.method = “Nelder-Mead”, emplik.method = “Owen”, optim. hessian = FALSE

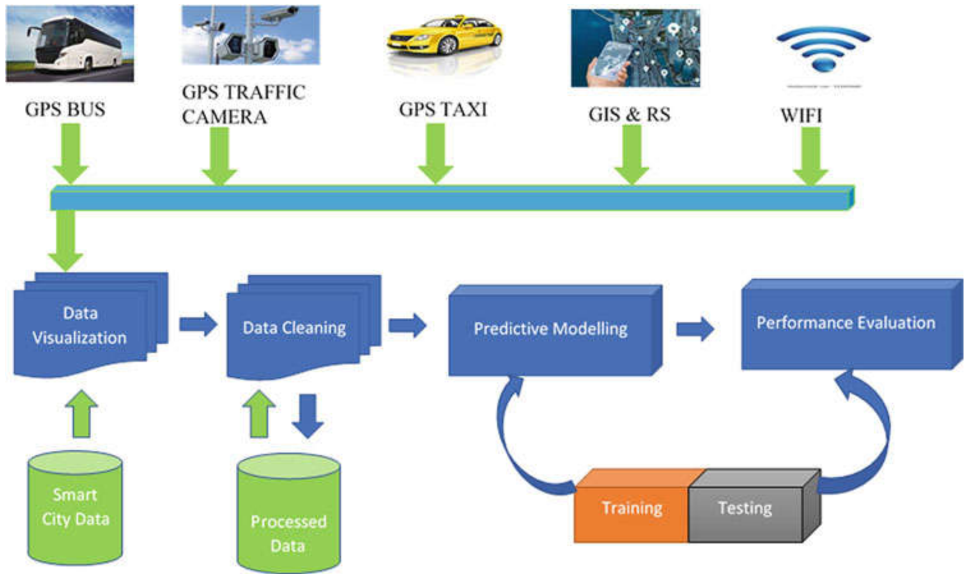

Fig. 1 Block diagram of smart city traffic pattern

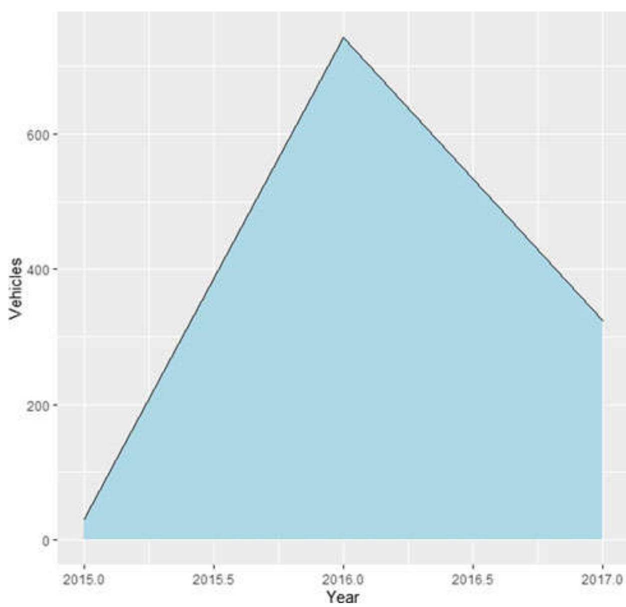

Fig. 2 Number of vehicles versus year

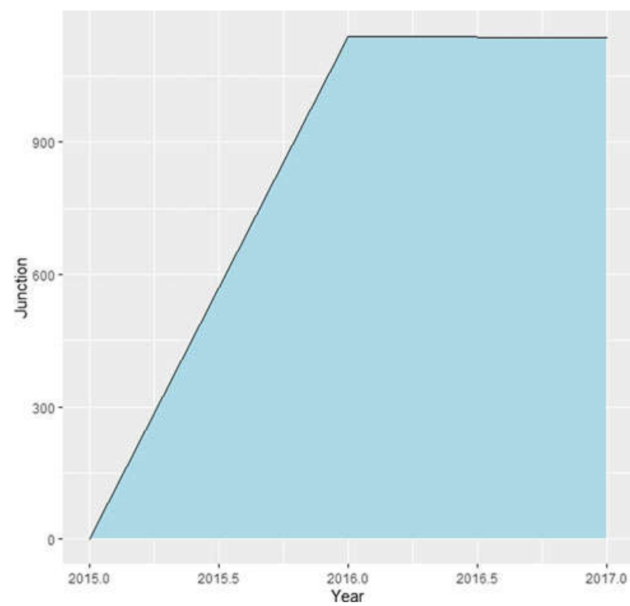

Fig. 3 Number of junctions versus year

Depicted in Fig. 7 are the output results of Actual (Vehicles) that are plying the roads in the city and the predicted values of the number of vehicles in 2017 using BAG, KNN, MARS, BGLM, and GLM models. Figures 8, 9, and 10 are containing the vehicular movements during times of the day such as morning, noon, afternoon, and evening for the year 2015, 2016, and 2017 respectively. There are more vehicular movement in the morning than afternoon, noon and evening for 2015. However, for 2016 and 2017, more vehicular movements are recorded in the afternoon than during any other time of the day.

In this research, we consider five machine learning algorithms to be specific: Bagging (BAG), K-Nearest Neighbor (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM). Figure 7 shows the machine learning algorithms utilized in this paper. Figure 6 shows the actual value of the vehicles and the predicted value of the machine learning such as BAG, KNN, MARS, BGLM, and GLM. Traffic congestion varies within the day event may cause surges in traffic at unexpected times. In Figs. 8, 9 and 10, we considered Morning, Afternoon,

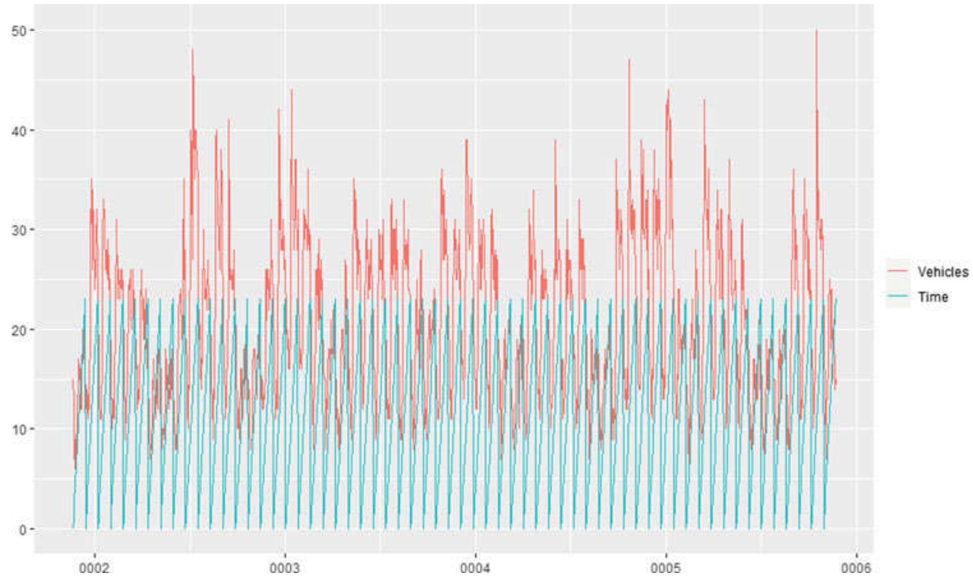

Fig. 4 Vehicles versus time in 2015

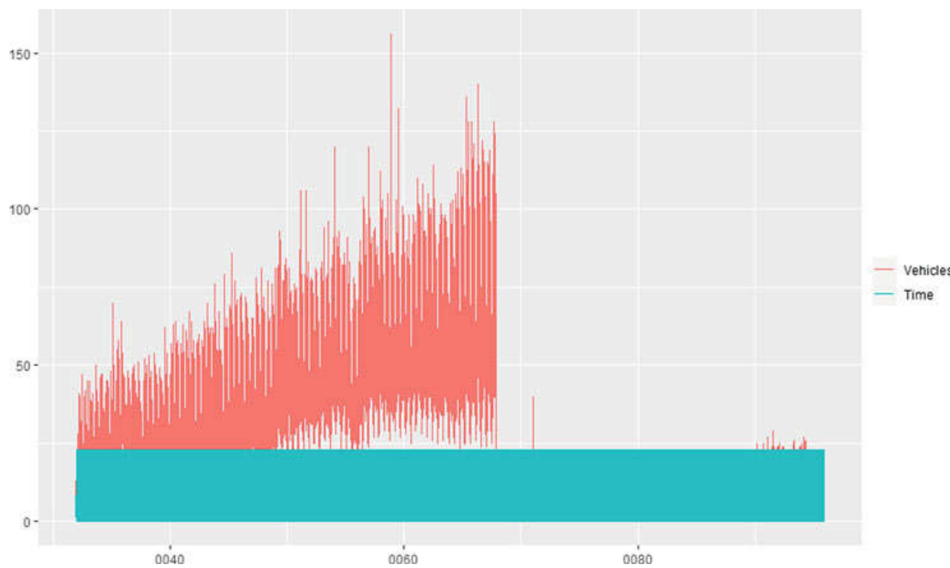

Fig. 5 Vehicles versus time in 2016

Evening, and Noon. The study revealed that in 2015 there is traffic surge in the Morning followed by Afternoon as shown in Fig. 8. In 2017 the traffic congestion increased exponentially which results in traffic surge in Afternoon than in the Noon as shown in Fig. 9. Depicted in Fig. 10 is the vehicular movement in 2017 which shows that Afternoon has the highest number of vehicle plying the road followed by Noon. The performance of each of the machine learning is shown in Table 4 using Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Scaled Error (MASE). GLM performs better than three the remaining four algorithms such as BAG, KNN, MARS, and BGLM having the least accuracy of 8.6802, 75.3468, 1.0498 using RMSE, MSE, MASE. From Table 3, we can deduce that GLM is a better predictor compared to the other models used in this work. Therefore, it is a promising algorithm for predicting vehicular movements.

5 Conclusion

As a new type of urban development strategy based on information technology, smart city building is committed to reforming the conventional urban organizing and management model to enable efficient data and resources distribution and solve urbanization problems. This research studied

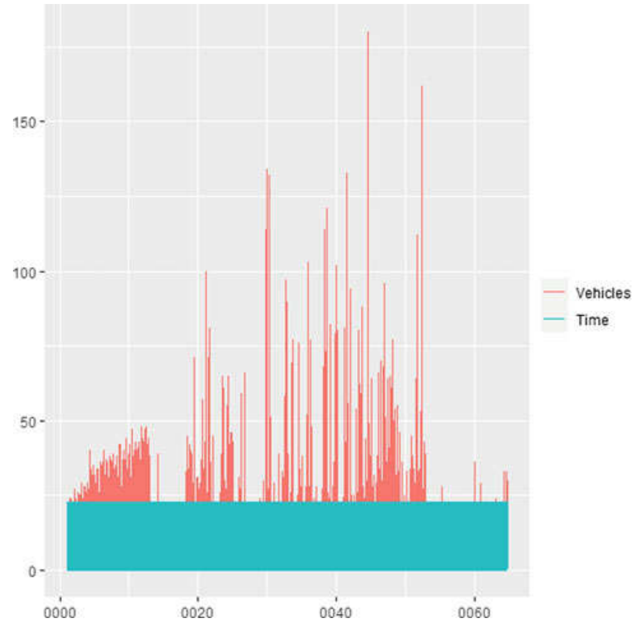

Fig. 6 Vehicles versus time in 2017

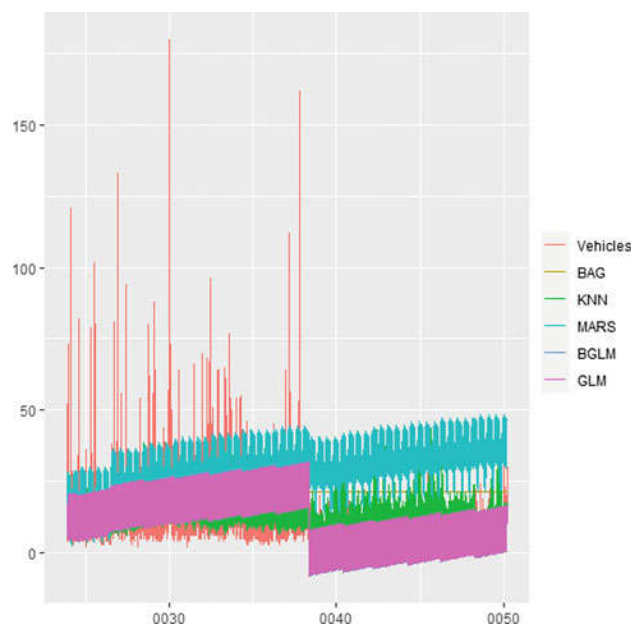

Fig. 7 Output results of actual (Vehicles) and predicted (BAG, KNN, MARS, BGLM, and GLM)

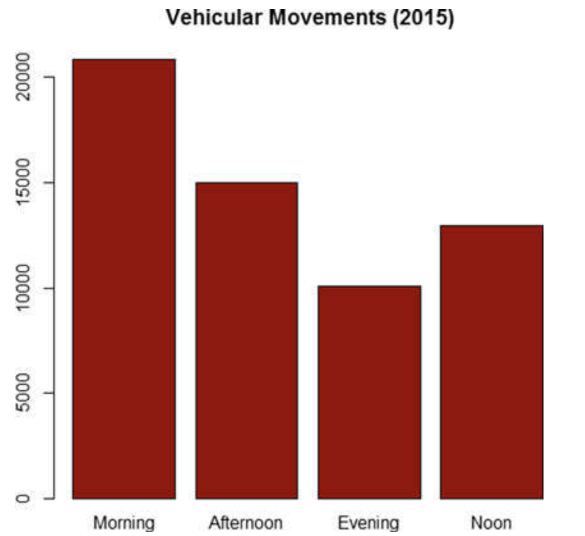

Fig. 8 Vehicular movements in 2015 during the day (Morning, Afternoon, Evening, Noon)

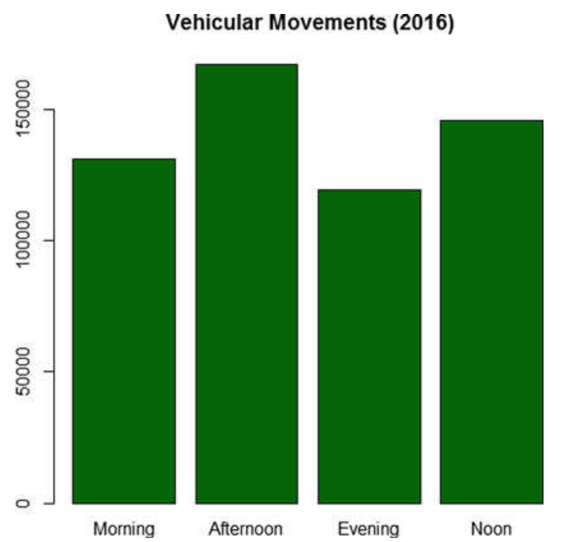

Fig. 9 Vehicular movements in 2016 during the day (Morning, Afternoon, Evening, Noon)

the transport congestion of the smart city. To address this gap in literature, this paper uses data from Kaggle database with 48,120 records from 2015 to 2017 to check if intelligent city-buildings such as road junction can mitigate congestion in cities. Results show that the building of smart cities such as road junctions greatly decreases the congestion level of urban traffic and increases urban efficiency. The mechanism check shows that the development of smart cities and information technology is a way to handle road issues but also to reduce the congestion of urban roads by urban overall creativity. In order to promote the development of smart cities and to alleviate congestions in city traffic, this research paper has significant practical implications. The limitations of the work include the inability of the dataset to indicate the number of buses, taxis, and traffic cameras on each junction which will enable us to determine the rate of vehicular queuing on each junction of the road.

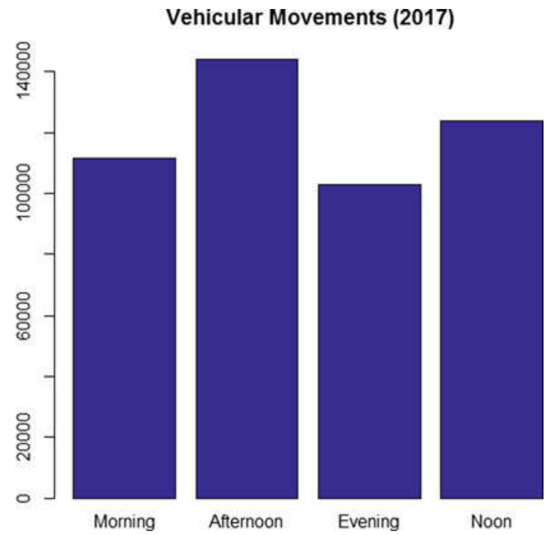

Fig. 10 Vehicular movements in 2017 during the day (Morning, Afternoon, Evening, Noon)

Table 4 Performance metrics of smart city traffic patterns

MODEL	RMSE	MSE	MASE
BAG	13.09740	171.5419	1.958938
KNN	9.236216	85.30769	1.093255
MARS	23.34668	545.0675	3.775596
BGLM	8.703736	75.75501	1.054976
GLM	8.680255	75.34683	1.049848

References

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method. *Applied Computing and Informatics, 12*(1), 90–108.

An, C., & Wu, C. (2020). Traffic big data assisted V2X communications toward smart transportation. *Wireless Networks, 26*(3), 1601–1610.

Bhattacharya, S., Somayaji, S. R. K., Gadekallu, T. R., Alazab, M., Maddikunta, P. K. R. (2020) A review on deep learning for future smart cities. *Internet Technology Letters*, e187.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1.

Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Reddy Maddikunta, P. K., Ra, I. H., & Alazab, M. (2020). Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics, 9*(2), 274.

Gosh, U. (2018). Smart city traffic pattern. https://www.kaggle.com/utathya/smart-city-traffic-patterns

Hinsbergen, C., Lint, J. W. C., & Sanders, F. (2007). Short term traffic prediction models. In *14th world congress on intelligent transport systems*.

Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft, C. H. (2019). Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers order in crumpled sheets. *Science Advances*, *5*(4), eaau6792.

Iwendi, C., Maddikunta, P. K. R., Gadekallu, T. R., Lakshmanna, K., Bashir, A. K., & Piran, M. J. (2020). A metaheuristic optimization approach for energy efficiency in the IoT networks. *Software: Practice and Experience*

Jia, Y., Wu, J., & Du, Y. (2016). Traffic speed prediction using deep learning method. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (pp. 1217–1222). IEEE.

Kumar, M. E., Reddy, G. T., Sudheer, K., Reddy, M. P. K., Kaluri, R., Rajput, D. S., & Lakshmanna, K. (2017). November. Vehicle theft identification and intimation using gsm and IoT. In *IOP conference series: Materials science and engineering* (Vol. 263, No. 4, p. 042062). IOP Publishing.

Morales, F., Ruiz, M., Gifre, L., Contreras, L. M., Lopez, V., & Velasco, L. (2017). Virtual network topology adaptability based on data analytics for traffic prediction. *IEEE/OSA J Opt Commun Networking*, *9*(1), A35–A45

Nallaperuma, D., Nawaratne, R., Bandaragoda, T., Adikari, A., Nguyen, S., Kempitiya, T., De Silva, D., Alahakoon, D., & Pothuhera, D. (2019). Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Transactions on Intelligent Transportation Systems, 20*(12), 4679–4690.

Ng, J. R., Wong, J. S., Goh, V. T., Yap, W. J., Yap, T. T. V., Ng, H. (2019). Identification of road surface conditions using IOT sensors and machine learning. In *Computational science and technology* (pp. 259–268). Springer.

Numan, M., Subhan, F., Khan, W. Z., Hakak, S., Haider, S., Reddy, G. T., Jolfaei, A., & Alazab, M. (2020). A systematic review on clone node detection in static wireless sensor networks. *IEEE Access, 8*, 65450–65461.

Oyewola, D. O., Dada, E. G., Omotehinwa, O. T., & Ibrahim, I. A. (2019). Comparative analysis of linear, non linear and ensemble machine learning algorithms for credit worthiness of consumers. *Computational Intelligence and Wireless Sensor Networks, 1*(1), 1–11.

Ozbayoglu, M., Kucukayan, G., & Dogdu, E. (2016). A real-time autonomous highway accident detection model based on big data processing and computational intelligence. In *2016 IEEE international conference on big data (Big Data)* (pp. 1807–1813). IEEE.

Raghavan, R., Singh, J. K., Reddy, T. G., Sudheer, K., Venkatesh, P., & Olabiyisi, S. O. (2017). A case study: Home environment monitoring system using internet of things. *International Journal of Mechanica Engineering of Technology, 8*(11), 173–180.

Reddy, G. T., Kaluri, R., Reddy, P. K., Lakshmanna, K., Koppu, S., Rajput, D. S. (2019). A novel approach for home surveillance system using IoT adaptive security. In *Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM)*. Amity University Rajasthan.

Reddy, T., Rm, S. P., Parimala, M., Chowdhary, C. L., Hakak, S., & Khan, W.Z. (2020). A deep neural networks based model for uninterrupted marine environment monitoring. *Computer Communications*, *157*, 64–75.

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of dimensionality reduction techniques on big data. *IEEE Access, 8*, 54776–54788.

Ruhlandt, R. W. S. (2018). The governance of smart cities: A systematic literature review. *Cities, 81*, 1–23.

Samadi, M., Afshar, M. H., Jabbari, E., & Sarkardeh, H. (2020). Application of multivariate adaptive regression splines and classification and regression trees to estimate wave-induced scour depth around pile groups. *Iranian Journal of Science and Technology, Transactions of Civil Engineering, 44*(1), 447–459.

Shi, G., Lim, C. Y., & Maiti, T. (2019). Bayesian model selection for generalized linear models using non-local priors. *Computational Statistics and Data Analysis, 133*, 285–296.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software, 39*(5), 1.

Verma, S., & Badade, S. (2019). Traffic prediction using machine learning. In *Proceeding of national conference on machine learning*.

Vinayakumar, R., Alazab, M., Soman, K., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access., 7*, 41525–41550.

Williams, N., Zander, S., & Armitage, G. J. (2006). A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *Computer Communication ACM SIGCOMM, 36*(5), 1–15.

Wu, Q., Huang, C., Wang, S. Y., Chiu, W. C., & Chen, T. (2007). Robust parking space detection considering inter-space correlation. In *2007 IEEE international conference on multimedia and expo* (pp. 659–662). IEEE.

Yang, J., Han, Y., Wang, Y., Jiang, B., Lv, Z., & Song, H. (2020). Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city. *Future Generation Computer Systems, 108*, 976–986.

Zheng, W., Lee, D. G., Asce, M., & Shi, Q. (2006). Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *Journal of Transportation Engineering, 132*(2), 1–8.