

Original Article

A Dynamic Model for Bus Arrival Time Estimation based on Spatial Patterns using Machine Learning

B. P. Ashwini¹, R. Sumathi², H. S. Sudhira³

^{1,2}Department of Computer Science and Engineering, Siddaganga Institute of Technology, Karnataka, India

³ Gubbi labs, LLP, Karnataka, India

¹Corresponding Author : ashvinibp@sit.ac.in

Received: 21 June 2022

Revised: 16 August 2022

Accepted: 11 September 2022

Published: 17 September 2022

Abstract - The notion of smart cities is being adapted globally to provide a better quality of living. A smart city's smart mobility component focuses on providing smooth and safe commuting for its residents and promotes eco-friendly and sustainable alternatives such as public transit (bus). Among several smart applications, a system that provides up-to-the-minute information like bus arrival, travel duration, schedule, etc., improves the reliability of public transit services. Still, this application needs live information on traffic flow, accidents, events, and the location of the buses. Most cities lack the infrastructure to provide these data. In this context, a bus arrival prediction model is proposed for forecasting the arrival time using limited data sets. The location data of public transit buses and spatial characteristics are used for the study. One of the routes of Tumakuru city service, Tumakuru, India, is selected and divided into two spatial patterns: sections with intersections and sections without intersections. The machine learning model XGBoost is modeled for both spatial patterns individually. A model to dynamically predict bus arrival time is developed using the preceding trip information and the machine learning model to estimate the arrival time at a downstream bus stop. The performance of models is compared based on the R-squared values of the predictions made, and the proposed model established superior results. It is suggested to predict bus arrival in the study area. The proposed model can also be extended to other similar cities with limited traffic-related infrastructure.

Keywords - Bus Arrival Time, Intelligent Transportation System, Machine Learning, Public Transit, Smart Cities.

1. Introduction

The notion of smart cities [1] [2] has facilitated institutions to exploit the potential of ICT in transforming cities to provide a better quality of life. The smart cities' smart mobility component [3] [4] primarily focuses on ensuring smooth and safe commuting between any two locations for the citizens. The mobility component also promotes the public transit system [5] as an eco-friendly alternative for private modes and emphasizes using ICT [6] in this direction. The recent developments in the Intelligent Transportation System (ITS) [7] [8] [9], Advanced Transportation System(ATS), and Advanced Passenger Information System (APIS) [10] aim at providing better LoS to commuters. One of the aims of the APIS is to assist the commuters by facilitating up-to-the-minute information about the public transit system and aid the commuters in taking decisions on their mode, route, etc.,

Several existing works on the Bus Travel and Arrival Time (BTAT) prediction are based on statistical, dynamic, time series, machine learning, and deep learning models [11]. Conventionally, travel time prediction is done on two scales [12]; long-term, where the travel time is predicted 1-2 days ahead of the current time, and short-term, where predictions

are made in 1-2 hours from the current time. The precision of the predictions mainly depends on the data sources available [13]; the static data such as road geometry, bus stops and intersections count, etc., along with the dynamic traffic flow information in the network plays a vital role. The availability of such datasets drives the accuracy of the prediction models. Several cities lack the infrastructure to capture live traffic flow information across the entire network. Installing the required infrastructure is mammoth and economically not viable as various past research has relied on the automatic vehicle location of buses and other vehicles, probe vehicle data, RFID at fixed locations, etc.

In this context, this work proposes a novel hybrid model to predict the Bus Arrival Time (BAT) in the short term using only the location data of the public transit buses. The proposed model combines the predictions of the extreme gradient boosting (XGBoost) machine learning technique and an analytical model developed using the preceding Bus Travel Time (BTT) and the spatial characteristics of the sections. The location data of the city service buses of Tumakuru is used for the study.



The remaining parts of the article are presented as follows. The existing literature on BTAT is reviewed and discussed in section 2. The proposed model for BTAT prediction using the limited dataset is illustrated in section 3. In section 4, the results and the discussions are presented. Finally, in section 5, the work is summarized as a conclusion.

2. Literature review

The recent developments in ICT and computing technology have leveraged the available resources to improve the quality of living across the globe. Travel/bus arrival time prediction is an important topic researched in the recent past, where the location data of the public transit buses are the most common dataset [13] used.

Authors [14] propose an approach to combine a radial basis neural network model and an online adjustment model. The online adjustment model adjusts the network model results based on the traffic's dynamic situation. To prove the validity, the model was deployed on a study route in Dalian city of China. The proposed model outperformed individual models such as linear regression, backpropagation, and radial basis function network models. In [15], a comparison of models such as historic average, Artificial Neural Network (ANN), and Kalman Filter(KF) for predicting the BTT is conducted. Authors considered robustness and accuracy of predictions as the metrics to compare the models and found that ANN outperforms the other models. The authors also suggested that these models can be used to implement the Advanced Public Transport System. Travel time prediction for multiple bus routes is proposed by [16] for routes in Shenzhen, China. The authors propose a hybrid model combining the baseline predictions of Support Vector Machines with the dynamic model KF.

Authors [17] have proposed a model to predict the travel time of buses in the short term in a dynamic manner using the historical data and the current information of the vehicle for which the prediction is made. A clustering algorithm: k-medoids, is used to extract the profiles in the historical data. The authors suggest that the proposed model may be used in the context of intercity trips. A performance comparison of the k-NN, KF, and Autoregressive Integrated Moving Average (ARIMA) models for dedicated and mixed traffic lanes is conducted [18]. The authors compared the model's individual and ensemble in predicting the arrival and travel time of buses in three cities in India. The dedicated lane route showed stability in prediction across models, whereas the combination of the mixed lanes k-NN-KF technique demonstrated the best results. A speed-based travel/arrival time prediction is proposed in [19]. The experiments were conducted in Samara, Russia. The proposed model was compared with Linear Regression (LR), basic ANN models, and Support Vector Regression (SVR), to highlight that it

can be used for the implementation of prediction systems in the future.

Similarly, authors in [20] have predicted the arrival time of buses using limited datasets, compared the LR and SVR models, and found that SVR performs better than LR. Authors in [21] have studied a route in Chennai, India, with high variance and have applied SVR to predict the BTT. The spatial-temporal SVRs developed demonstrated improvement in the predictions compared to the existing model deployed. In a study [22] conducted in Charlotte, North Carolina, travel times are predicted for short horizons analyzed the performance of four machine learning models Regression Trees, XGBoost, Long Short-Term Memory (LSTM) neural network, and Random Forests (RF). The performance was assessed based on prediction accuracy and reliability, and RF showed better performance in both accuracy and reliability. Authors in [23] have highlighted the importance of a deep understanding of the spatial-temporal features through Recurrent Neural Networks. They have proposed a wide and deep model for travel time prediction. The experiments were conducted using the datasets of two big cities in China, and the model proposed demonstrated good results. Authors in [24] have compared the suitability of linear and non-linear machine learning models for BTT prediction. They have concluded that for tier-2 cities like Tumakuru, non-linear models are suitable, and ensemble tree models such as RF and Gradient Boosting Regression Trees(GBRT) are best for schedule preparation and travel time prediction.

Overall, it is concluded from the existing literature that there are no generic models for accurate BTAT prediction, and most of the solutions are location specific. With the advancements in ICT and ITS, there are several cities in which these technologies are yet to be implemented. In this direction, this work aims to predict BTAT in the study area with limited traffic-related infrastructure. Hybrid and machine learning models have been widely used in BTAT recently; hence a suitable machine learning model is identified from the previous study, and a hybrid model is proposed.

3. Data

The location data of the Tumakuru city service buses of March 2021 is used to develop the machine learning and bus arrival estimation model. 600 trips of the route, Tumkur Bus stand to Kyathasandra is selected; also, the other routes sharing the common sections with the study route are identified for modeling. The sample location data is given in table 1, and the study route map with the bus stops is given in Figure 1. The information on the study route augmented with the spatial characteristics is given in table 2.

Table 1. Sample location data of public transit bus

Vehicle No	Date and Time	Latitude	Longitude	Odometer	Speed
KA-06-F-0836	01-03-2021 07:33:37	13.34286	77.09886	6121.51	0
KA-06-F-0836	01-03-2021 07:33:44	13.34292	77.09894	6121.52	6
KA-06-F-0836	01-03-2021 07:33:54	13.34298	77.09917	6121.54	6
KA-06-F-0836	01-03-2021 07:34:04	13.34314	77.09919	6121.56	7
KA-06-F-0836	01-03-2021 07:34:14	13.34328	77.09897	6121.6	12
KA-06-F-0836	01-03-2021 07:34:24	13.34313	77.09872	6121.63	12
KA-06-F-0836	01-03-2021 07:34:34	13.34271	77.09856	6121.68	21
KA-06-F-0836	01-03-2021 07:34:44	13.3422	77.09839	6121.74	17
KA-06-F-0836	01-03-2021 07:34:54	13.3419	77.09829	6121.78	0
KA-06-F-0836	01-03-2021 07:34:58	13.34181	77.09826	6121.79	8
KA-06-F-0836	01-03-2021 07:35:08	13.34144	77.09814	6121.83	19
KA-06-F-0836	01-03-2021 07:35:18	13.34085	77.09798	6121.9	26
KA-06-F-0836	01-03-2021 07:35:28	13.3403	77.09781	6121.96	17
KA-06-F-0836	01-03-2021 07:35:38	13.33983	77.09767	6122.01	22
KA-06-F-0836	01-03-2021 07:35:48	13.33932	77.09753	6122.07	15
KA-06-F-0836	01-03-2021 07:35:58	13.33908	77.09748	6122.1	9

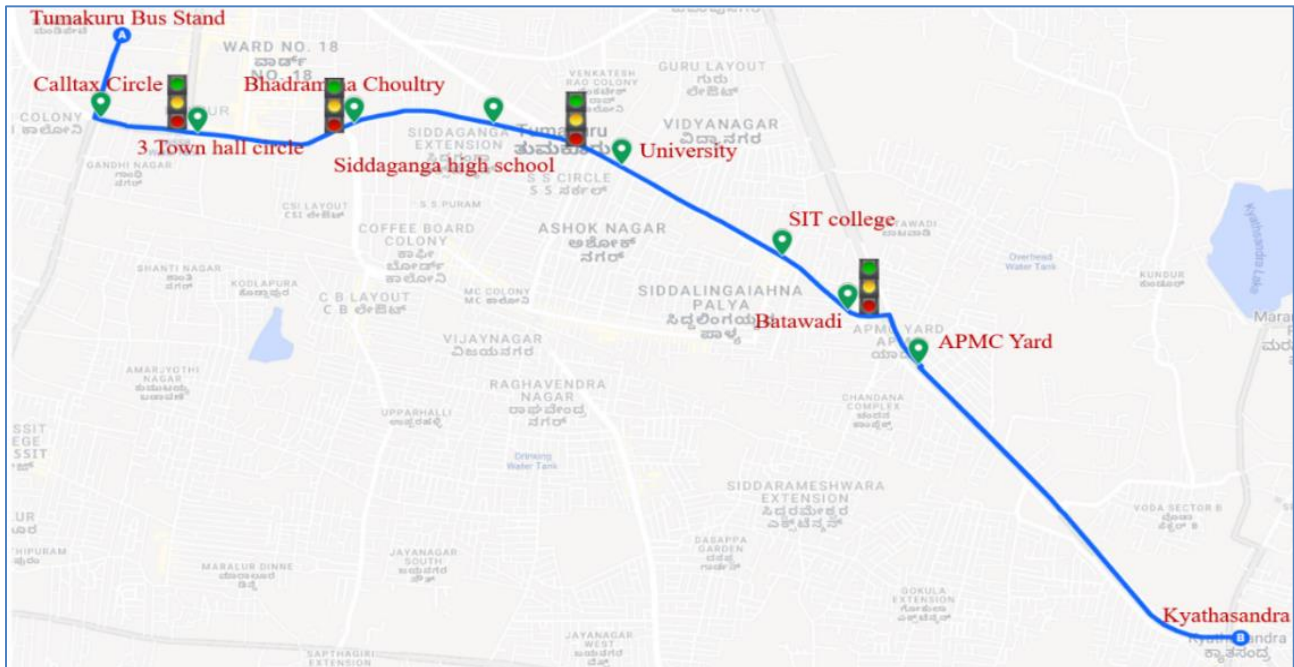


Fig. 1 The route map of the Tumkur Bus stand to Kyathasandra with bus stops

Table 2. Route information

Section	Start and end bus stops	Length (meters)	LUP	Intersection
1	Tumkur Bus Stand(TBS) - CallTax Circle (CTC)	600	CBD	No
2	CallTax Circle – Town Hall Circle (THC)	450	CBD	Yes
3	Town Hall Circle - Bhadranna Choultry (BC)	750	CBD	Yes
4	Bhadranna Choultry - Siddagnaga High School (SHS)	650	IC	No
5	Siddagnaga High School – Government University (GU)	550	IC	Yes
6	University- SIT College (SIT)	1000	ISU	No
7	SIT College – Batawadi (BT)	350	ISU	Yes
8	Batawadi - APMC Yard (APMC)	400	OSU	No
9	APMC Yard – Kyathasandra (KY)	2200	OSU	No

CBD: Central Business District, IC: Inner City, ISU: Inner Suburban, OSU: Outer Suburban

4. Methods

4.1 Preprocessing

Variability analysis of the travel times in the selected route (up and downstream) is presented in [25]. The same procedure is followed to pre-process the data in the current article. The raw location data are pre-processed to remove missing and erroneous entries and are transformed into aggregates at the bus stop level for further study. The travel time between bus stops depends on factors such as Land Use patterns (LUP), signalized intersections, un-signalized intersections, section length, etc. As an additional feature, LUP and the length of the section are augmented to the trip aggregates. Unsignalized intersections are not included in the study. The presence of signalized intersections has a major influence on travel time; hence the trip aggregates are split as sections (bus stop to bus stop) with and without intersections. Sections 2,3,5, and 7 are Sections with InterSections (SIS) and sections 1,4,6,8 and 9 are the sections without intersections and are called Normal Sections (NS) in the future in this article.

4.2. Extreme Gradient Boosting (XGBoost)

XGBoost is a tree model used in applications that need classification and regression. It is an efficient implementation of the gradient boosting regression tree technique with boosting approach that ensembles multiple simple regression trees. The trees are built based on the residuals at each level to optimize the loss function. It is a widely used model to predict BTT [26][27][28]. The scikit-learn [29], a free machine learning library, is used to implement XGBoost. According to the previous study [24], the non-linear models are best suited for BTT prediction in the current study area. Therefore, the XGBoost technique is used to forecast the BTT. The XGBoost is trained for trip aggregates data with intersections and without intersections separately, and individual models are developed. A few parameters and its corresponding values in the study are given in table 3.

4.3. Bus Arrival Time Estimation Model

The real-time information on the traffic flow in the link, congestion alerts, and occurrence of incidents are unavailable

in the study area. The data available is limited to only the location data of the buses. Hence, the preceding buses are considered probe vehicles, and the travel time of the preceding buses is assumed to reflect the delays in the link. The latest travel time information about the sections, such as travel time and running speed, are extracted from preceding buses of the same and other routes (multi-routes). The trips that have preceding trips with less than 30 minutes difference from the current trip start time are considered for the study. Based on the spatial patterns, the sections of the route are divided as follows:

- Sections without intersections (Normal Section (NS))
- Sections with intersections (SIS)

A dynamic BAT Estimation Model (BATEM) is proposed in this work that predicts the BAT at a downstream bus stop based on the preceding trip data of buses and the machine learning model predictions for both spatial patterns. The implementation flow of the proposed model is given in figure 2. The list of symbols used in the article is summarized in table 4.

The proposed BATEM estimates the BAT for individual bus stops by considering the current time as the start time of the section. The BAT for each section j is estimated as given in equation 1, in which C_{time} is the current time and the ATT_t^j is the Adjusted Travel Time estimated for section j at start time t .

$$BAT_j = C_{time} + ATT_t^j \tag{1}$$

Adjusted Travel Time is the time estimated based on the type of the section using equation (2). It is the weighted average of the Forecasted Travel Time FTT_t^j by the XGBoost, and the estimated travel time based on preceding trip travel time PTT_i^j . The weights are computed as given in equation (5).

Table 3. Few important parameters of the XGBoost Regression technique used

Parameters	Values Used	Description
objective	squared error	A loss function that estimates the distance of a predicted value from the actual regressed value
alpha	1.00	L1 regularization term on weights
learning_rate	0.05	Step size used in each iteration during optimizing a loss function
n_estimators	200	Number of trees in the additive model
colsample_bytree	0.60	A subsample ratio of columns during the construction of each tree. Subsampling is done for every constructed tree
max_depth	3.00	The upper bound for the number of nodes in the weak learner regression tree

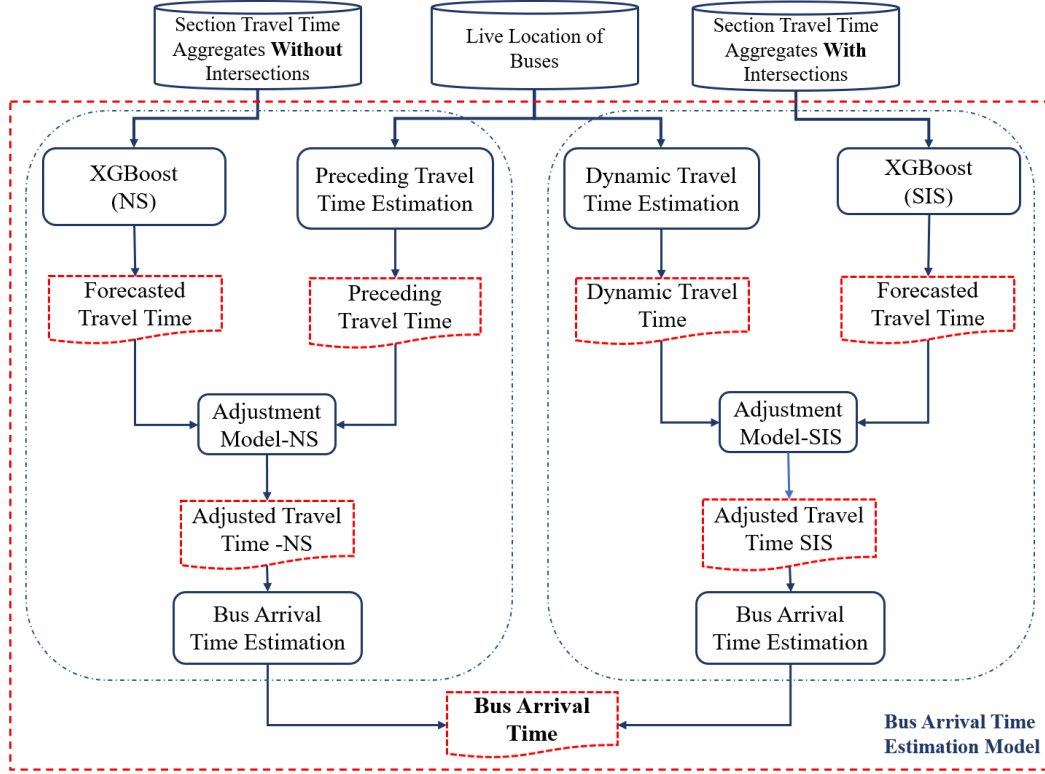


Fig. 2 Flow of the proposed model

$$ATT_t^j = \begin{pmatrix} w_1 * FTT_t^j + w_2 * PTT_i^j & \text{for NS} \\ w_1 * FTT_t^j + w_2 * DTT_i^j & \text{for SIS} \end{pmatrix} \quad (2)$$

For normal sections, the ATT_t^j is estimated using the FTT_t^j and preceding trip travel time PTT_i^j , whereas for sections with intersections, the estimations are made using Dynamic Travel Time DTT_i^j and FTT_t^j . The PTT_i^j is the travel time of a preceding bus for the section with start time i , where $(t - i)$ is less than 30 minutes. If multiple bus information is available, then the minimum of $(t - i)$ is considered. and the DTT_i^j is estimated using the running time of the preceding trip PRT_i^j Along with the standard dwell time DWD_j And average delay at intersections ID_j as given in equation (3).

$$DTT_i^j = PRT_i^j + DWD_j + ID_j \quad (3)$$

The equation to estimate the preceding trip running time is presented in equation (4), where d_j is the distance of section j , and RS_i^j is the running speed of each section of the preceding trip.

$$PRT_i^j = \frac{d_j}{RS_i^j} \quad (4)$$

The weights in equation (2) are estimated by equation (5). These weights are derived by estimating the x_1 ; the coefficient of determination (R-squared) of the machine learning model and x_2 The historical data shows the correlation of the current trip to that of the preceding trip.

$$w_1 = \frac{x_1}{x_1 + x_2} \quad w_2 = \frac{x_2}{x_1 + x_2} \quad (5)$$

The R-squared (R2) is a statistical measure determining the strength of the linear relationship between the parameters under consideration. The formula for computing the coefficient of determination is given in equation 6 for n trips. Where $p_{y,j}$ is the dependent variable (actual travel time of trip y for section j), the function $f(T_{z,j})$ returns the predicted travel time using the independent variables, the $T_{z,j}$ is a table containing the independent variables day-of-the-week, the trip's start time, section number, and the land use pattern. $\mu p_{y,j}$ is the mean of the actual travel time in test data.

$$x_1 = 1 - \frac{\sum_{y=1}^n (p_{y,j} - f(T_{z,j}))^2}{\sum_{i=1}^n (p_{y,j} - \mu p_{y,j})^2} \quad (6)$$

Analyzing the correlation [30] is a method used to measure the level of deviation in one feature due to deviation in another. The correlation analysis of the preceding trip travel and the current trip travel time in the archived data is

conducted to derive the correlation coefficients. Preceding trips traversing the same sections within 30 minutes to the current time are considered. The standard equation for computing the correlation is given in equation 7, where x_2 is the correlation coefficient, tt_c is the current trip travel time and tt_p is the travel time of the preceding trip. The μtt_c and μtt_p are the average of the current and preceding trip travel times.

$$x_2 = \frac{\Sigma((tt_c - \mu tt_c) * (tt_p - \mu tt_p))}{\sqrt{\Sigma(tt_c - \mu tt_c) * \Sigma(tt_p - \mu tt_p)}} \quad (7)$$

All the equations are applied to sections with intersections and without intersections to dynamically predict the BAT at a downstream bus stop using only the live and historic location data of the buses.

5. Results and Discussion

The traffic's stochastic behaviour sometimes decreases the machine learning model's performance. The live data regarding traffic flow, incidents, and delays influence the travel duration and, thus, the BAT. With limited information, such as only live and historical location data of the city service buses, predicting the arrival time of the buses is a colossal task. Hence, this article presents a hybrid mode of estimating the arrival time of the buses, where the machine learning model and a model based on the preceding trips are combined to predict the BTAT. The estimated values for the parameters in the proposed model are tabulated in table 5. The average intersection delay for each intersection is estimated in [34].

The R-squared (x_1) value indicates that the sections without intersections are more predictable by the machine learning model than those with intersections. The intersections in Tumakuru city operate in clearance mode with no bus priority [32] [33]. The length of sections with intersections is less than 800 meters. Thus, the variations in the waiting delay at signalized intersections affect the travel time. Hence the sections having intersections demonstrate more variations and less predictability.

A plot of the actual travel time against the forecasted travel by the XGBoost model and the travel time estimated based on the proposed model of a few sample trips is presented in Figure 3 and Figure 4. Figure 3 depicts the predictions of the sections with intersections, and figure 4 of the sections without intersections. The start time for each section and the travel time in seconds is plotted on the X-axis and the Y-axis, respectively. It is evident from the plot that the proposed model can recognize the dynamic changes in the traffic flow and estimate the travel time better as compared to the XGBoost model.

Table 5. The computed values of the constants used in the study

Spatial level	w ₁	w ₂	x ₁	x ₂	Intersection delay in seconds (section)
SIS	0.45	0.55	0.40	0.50	47 (2), 45 (3), 57 (5), 20 (7)
NS	0.56	0.44	0.71	0.55	-----

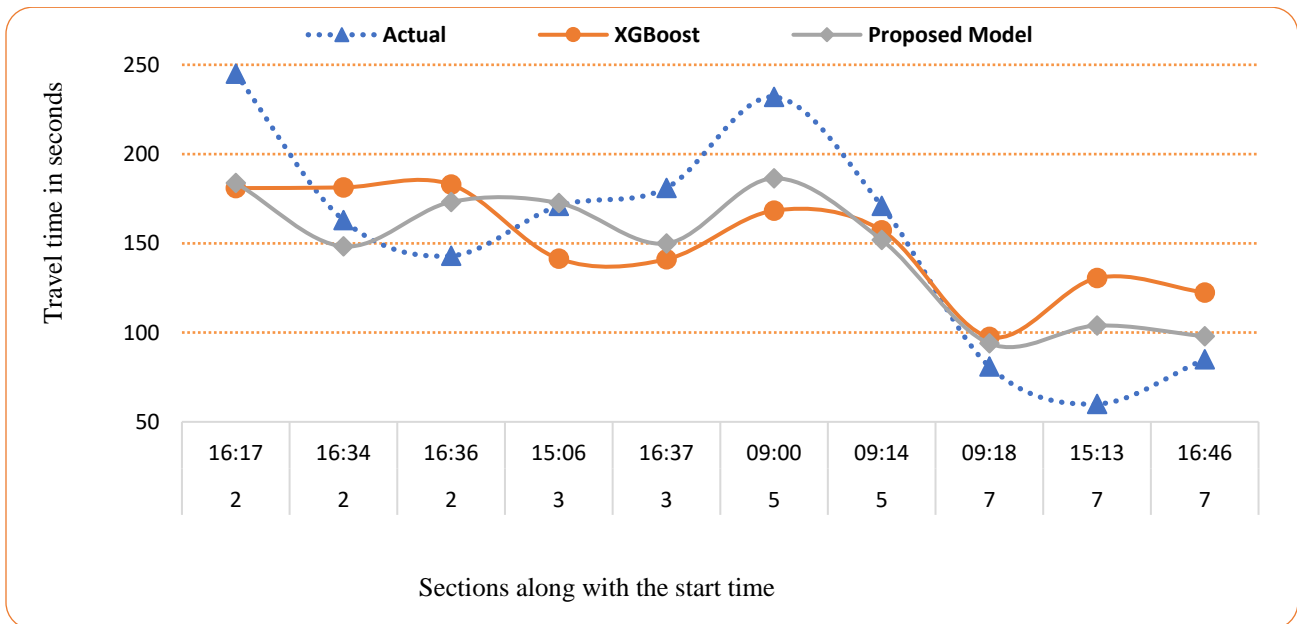


Fig. 3 Actual vs predicted travel times of XGBoost and proposed models for SIS

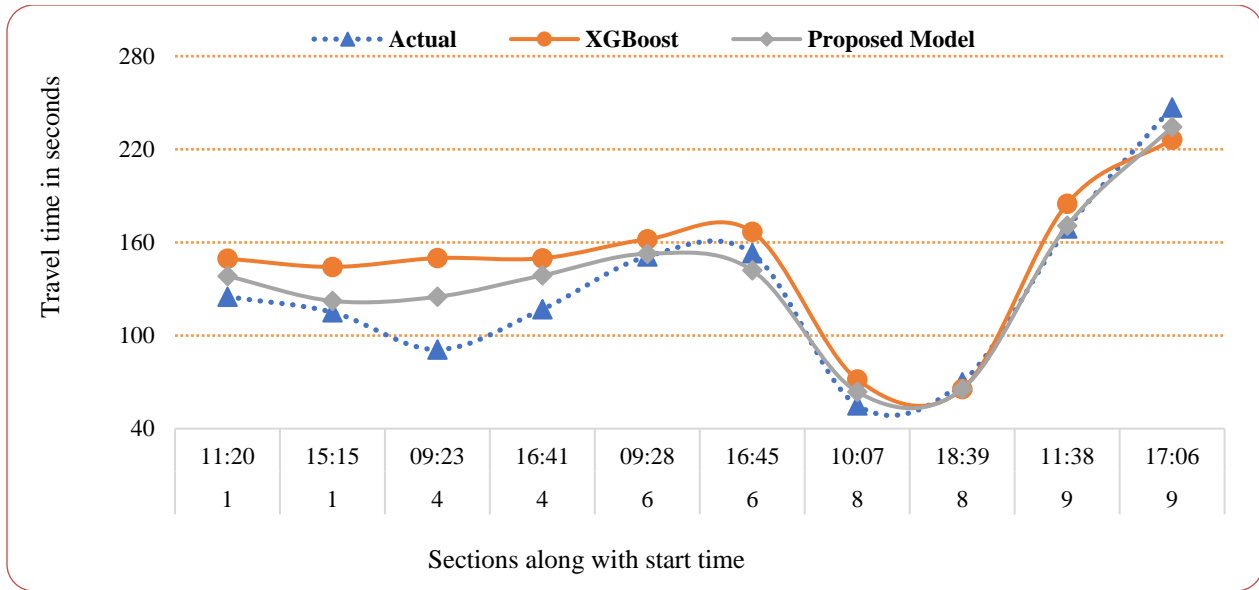


Fig. 4 Actual vs predicted travel times of XGBoost and proposed models for NS

The predictions made for one trip on the 11th of Marth 2021, starting at 15:14:00, are plotted to depict the predictions of the XGBoost and the proposed model in figure 5. The actual travel time of the trip is 1209 seconds, the total duration predicted by XGBoost is 1339 seconds, and the

proposed model is 1261. The XGBoost predicted individual sections' travel time with a mean error of 30 seconds, whereas the proposed models' average error was 18 seconds. It implies that the travel time predicted by the proposed model is better than the machine learning model.

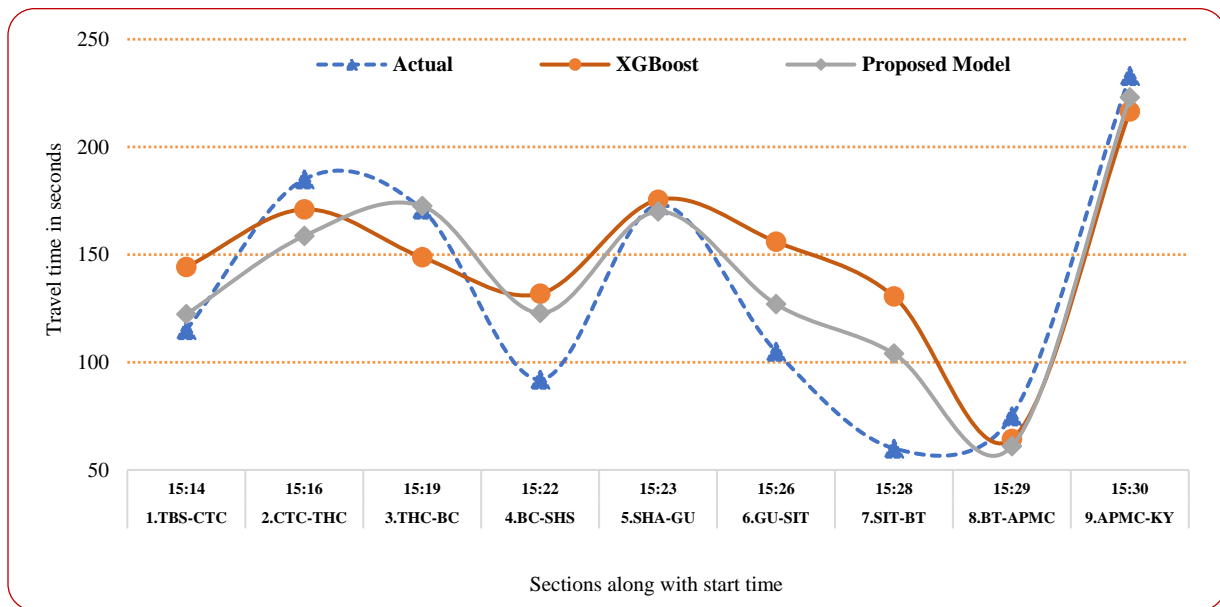


Fig. 5 Actual vs predicted travel times of XGBoost and proposed models for a sample trip

The R-squared of the XGBoost model is 0.4 and 0.71. In contrast, the R-squared of the proposed model is 0.59 and 0.80 for sections with and without intersections, respectively, demonstrating significant improvement in the predictions. The performance values are plotted in Fig. 6.

The data available in the study area is limited to the locations of the public transit buses. With such a constraint on the data, where the live information of the traffic, congestion, or incidents is unavailable, the proposed model has attempted to predict the BAT. The model's performance suggests that the proposed model is suitable for predicting

BAT in the study area. And there are several similar cities with limited traffic infrastructure for which the proposed model can be extended. To further improve the performance of the proposed model, the current trip information, dynamic traffic flow information of the links, and details of incidents, if any, are also vital as the up-to-the-minute information reduces the uncertainties of the prediction.

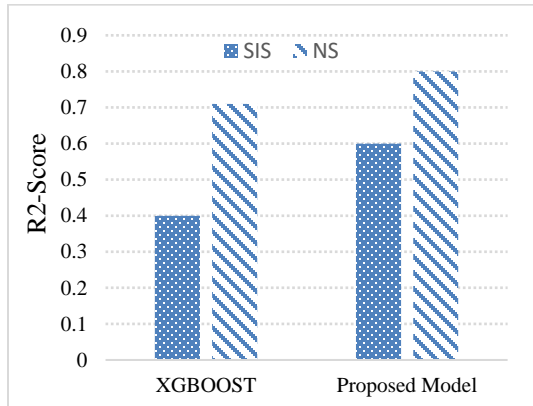


Fig. 6 Performance comparison of the proposed model with XGBOOST

6. Conclusion

A dynamic model to predict bus arrival time using only limited data sets based on machine learning is proposed in this article. Based on the previous study results, the base travel time is predicted using the XGBoost model. The

proposed Bus Arrival Time Estimation Model estimates the bus arrival time for individual bus stops for a selected route in Tumakuru, India. The sections between the bus stops were divided into two discrete spatial patterns: a section with intersections and sections without intersections. Each spatial pattern was modeled separately using the XGBoost and the proposed model. The proposed model estimated the bus arrival time based on both machine learning model results and the preceding trip travel time and demonstrated better accuracy than the XGBoost model for both spatial patterns. The preceding trip within 30 minutes to that of the current demonstrated a positive correlation, illustrating its usefulness in forecasting travel times. The proposed model exhibited better accuracy based on the R-squared values. The proposed model can be suggested for bus arrival time forecasting in the study route and extended to other routes to predict the city-wide bus arrivals. Similar cities can employ the proposed model as well. If available, the up-to-the-minute details of the traffic flow, incidents, and other delays can further improve the forecasting. The model can be updated in the future once the up-to-the-minute data stream is available. Overall, with the available limited datasets, the proposed model is a promising option for forecasting the bus arrival time.

Acknowledgements

The authors wish to thank Smart City Limited of Tumakuru (TSCL) for providing the required logs of the city service buses.

References

- [1] S. Joshi, S. Saxena, T. Godbole, En Shreya, "Developing Smart Cities: An Integrated Framework", *Procedia Comput. Sci.*, vol 93, no. September, pp. 902–909, 2016, Doi: 10.1016/J.Procs.2016.07.258.
- [2] M. S.Kumbhar En P. S.Yalagi, "Urban Resources for Smart City Application", *Int. J. Eng. Trends Technol.*, vol.40, no.6, pp. 366–370, 2016, Doi: 10.14445/22315381/Ijett-V40p259.
- [3] Polytechnique Montreal, "Smart Cities and Integrated Mobility: A White Paper", *Next Gener. Integr. Mobility Drive. Smart City*, pp. 1–78, 2018.
- [4] R. M. Savithamma, B. P. Ashwini, En R. Sumathi, "Smart Mobility Implementation in Smart Cities: A Comprehensive Review on State-of-Art Technologies", *2022 4th Int. Conf. Smart Syst. Inven. Technol.*, pp. 10–17, 2022, Doi: 10.1109/Icssit53264.2022.9716288.
- [5] M. Kumar, N. Kumari, S. tomar, En T. Kumar, "Smart Public Transportation for Smart Cities", *SSRN Electron. J.*, pp. 1–6, 2019, Doi: 10.2139/SSrn.3404487.
- [6] S. Porru, F. E. Misso, F. E. Pani, En C. Repetto, "Smart Mobility and Public Transport: Opportunities and Challenges in Rural and Urban Areas", *J. Traffic Transp. Eng. English Ed*, vol 7, no 1, pp. 88–97, 2020, Doi: 10.1016/J.Jtte.2019.10.002.
- [7] R. S. Chhillar, "A Review of Intelligent Transportation Systems in Existing Framework Using Iot", vol.70, no. 6, pp. 137–143, 2022.
- [8] S. Muthuramalingam, A. Bharathi, S. Rakesh Kumar, N. Gayathri, R. Sathiyaraj, En B. Balamurugan, "Iot Based Intelligent Transportation System (Iot-Its) for Global Perspective: A Case Study", *Intell. Syst. Ref. Libr.*, vol.154, pp. 279–300, 2019, Doi: 10.1007/978-3-030-04203-5_13.
- [9] K. Iqbal, M. A. Khan, S. Abbas, Z. Hasan, En A. Fatima, "Intelligent Transportation System (Its) for Smart-Cities Using Mamdani Fuzzy Inference System", *Int. J. Adv. Comput. Sci. Appl.*, vol.9, no.2, pp. 94–105, 2018, Doi: 10.14569/Ijacs.2018.090215.
- [10] A. Monzon, S. Hernandez, En R. Cascajo, "Quality of Bus Services Performance: Benefits of Real Time Passenger Information Systems", *Transp. Telecommun.*, vol.14, no.2, pp. 155–166, 2013, Doi: 10.2478/Tj-2013-0013.
- [11] M. Bai, Y. Lin, M. Ma, En P. Wang, *Travel-Time Prediction Methods: A Review*, vol.11344, Lncs. Springer International Publishing, 2018.

- [12] D. Panovski En T. Zaharia, "Long and Short-Term Bus Arrival Time Prediction With Traffic Density Matrix", *Ieee Access*, vol.8, pp. 226267–226284, 2020, Doi: 10.1109/Access.2020.3044173.
- [13] P. B. Ashwini En R. Sumathi, "Data Sources for Urban Traffic Prediction: A Review on Classification, Comparison and Technologies", *Proc. 3rd Int. Conf. Intell. Sustain. Syst. Iciss 2020*, pp. 628–635, 2020, Doi: 10.1109/Iciss49785.2020.9316096.
- [14] L. Wang, Z. Zuo, En J. Fu, "Bus Arrival Time Prediction Using Rbf Neural Networks Adjusted By online Data", *Procedia - Soc. Behav. Sci.*, vol 138, no.0, pp. 67–75, 2014, Doi: 10.1016/J.Sbspro.2014.07.182.
- [15] W. Fan En Z. Gurmu, "Dynamic Travel Time Prediction Models for Buses Using only Gps Data", *Int. J. Transp. Sci. Technol.*, vol.4, no.4, pp. 353–366, 2015, Doi: 10.1016/S2046-0430(16)30168-X.
- [16] C. Bai, Z. R. Peng, Q. C. Lu, En J. Sun, "Dynamic Bus Travel Time Prediction Models on Road With Multiple Bus Routes", *Comput. Intell. Neurosci.*, vol.2015, 2015, Doi: 10.1155/2015/432389.
- [17] G. De Blasio En C. R. Garc, "Bus Travel Time Prediction Model Based on Profile Similarity", *Encycl. Soc. Netw. Anal. Min.*, pp. 1942–1942, 2018, Doi: 10.1007/978-1-4939-7131-2_100921.
- [18] R. Jairam, B. A. Kumar, S. S. Arkatkar, En L. Vanajakshi, "Performance Comparison of Bus Travel Time Prediction Models Across Indian Cities", *Transp. Res. Rec.*, vol.2672, no.31, pp. 87–98, 2018, Doi: 10.1177/0361198118770175.
- [19] A. A. Agafonov En A. S. Yumaganov, "Performance Comparison of Machine Learning Methods in the Bus Arrival Time Prediction Problem", *Ceur Workshop Proc.*, vol.2416, pp. 57–62, 2019, Doi: 10.18287/1613-0073-2019-2416-57-62.
- [20] A. Kviesis, A. Zacepins, V. Komasilovs, En M. Munizaga, "Bus Arrival Time Prediction With Limited Data Set Using Regression Models", no Vehits 2018, pp. 643–647, 2019, Doi: 10.5220/0006816306430647.
- [21] A. K. Bachu, K. K. Reddy, En L. Vanajakshi, "Bus Travel Time Prediction Using Support Vector Machines for High Variance Conditions", vol.36, no.3, pp. 221–234, 2021.
- [22] B. Qiu En W. D. Fan, "Machine Learning Based Short-Term Travel Time Prediction : Numerical Results and Comparative Analyses," 2021.
- [23] Y. Yuan, C. Shao, Z. Cao, Z. He, En C. Zhu, "Bus Dynamic Travel Time Prediction : Using a Deep Feature Extraction Framework Based on Rnn and Dnn", *Electron, Mdpi*, 2020.
- [24] B. P. Ashwini, R. Sumathi, En H. S. Sudhira, "Bus Travel Time Prediction: A Comparative Study of Linear and Non-Linear Machine Learning Models", *J. Phys. Conf. Ser.*, vol.2161, no.1, pp.012053, 2022, Doi: 10.1088/1742-6596/2161/1/012053.
- [25] R. Ashwini, B.P., Sumathi, "Variability Analysis of Public Transit Bus Travel Time Using Gps Logs: A Case Study", in *Proceedings of the Third International Conference on Information Management and Machine Intelligence. Algorithms for Intelligent Systems. Springer, Singapore.*, pp. 567–577, 2022. Doi: 10.1007/978-981-19-2065-3_60.
- [26] F. Zhang, X. Zhu, T. Hu, W. Guo, C. Chen, En L. Liu, "Urban Link Travel Time Prediction Based on a Gradient Boosting Method Considering Spatiotemporal Correlations", *Isprs Int. J. Geo-Information*, vol.5, no.11, 2016, Doi: 10.3390/Ijgi5110201.
- [27] X. Li En R. Bai, "Freight Vehicle Travel Time Prediction Using Gradient Boosting Regression Tree", no.December, pp. 1010–1015, 2017, Doi: 10.1109/Icmla.2016.0182.
- [28] Y. Zhang En A. Haghani, "A Gradient Boosting Method to Improve Travel Time Prediction", *Transp. Res. Part C Emerg. Technol.*, vol.58, pp. 308–324, 2015, Doi: 10.1016/J.Trc.2015.02.019.
- [29] A. G. F Pedregosa, G Varoquaux, "Scikit-Learn: Machine Learning in Python", *J. Mach. Learn. Res.*, vol.12, pp. 2825–2830, 2011, [online]. Available At: <https://scikit-learn.org/dev/modules/ensemble.html#gradient-tree-boosting>.
- [30] S. Senthilnathan, "Usefulness of Correlation Analysis", *SSRN Electron. J.*, 2019. Doi: 10.2139/SSrn.3416918.
- [31] C. Managwu, D. Matthias, N. Nwaibu, "Random forest Regression Model for Estimation of Neonatal Levels in Nigeria," *SSRG International Journal of Computer Science and Engineering*, vol.7, no. 7, pp. 41-44, 2020. *Crossref*, <https://doi.org/10.14445/23488387/Ijcse-V7i7p107>
- [32] E. Summary, "Comprehensive Mobility Plan for Cma Final Report," no. January, 2014.
- [33] U. L. Transport, "Tumkur City Bus Service Directorate of Urban Land Transport," 2013, [online]. Available At: [http://www.urbantransport.kar.gov.in/Tumkur City Bus Service.Pdf](http://www.urbantransport.kar.gov.in/Tumkur%20City%20Bus%20Service.Pdf).
- [34] R. Savithramma, R.M., Sumathi, "Comparative Analysis of Waiting Delay Estimation Models for Signalized Intersections: A Case Study Tumakuru City," in *Proceedings of the Third International Conference on Information Management and Machine Intelligence, Algorithms for Intelligent Systems*, pp. 579–589, 2022. Doi: 10.1007/978-981-19-2065-3_61.