# Machine Learning-Driven Forecasting of Traffic Congestion

MD. SAKIB ALAM, Brac University, Bangladesh
TASFIQUE ZAMAN CHOWDHURY SIFAT, Brac University, Bangladesh
LAMISHA RAHMAN, Brac University, Bangladesh

**Abstract**

The ongoing urbanization has led to severe traffic congestion problems in the over populated cities like Dhaka which is affecting transportation time, fuel consumption, mental health, environmental impacts like decreasing air quality and reducing productivity in daily life by causing delay while reaching destination. The Intelligent Transportation System (ITS) has played a great role in developing cities to resolve traffic problems and introducing a better traffic management system. This paper aims to improve traffic congestion prediction in over crowded areas by analyzing data by Machine Learning based multi step prediction approach. Furthermore to provide a more optimized prediction method and find a better route to reach the destination in order to increase productivity. Forecasting traffic is mandatory to introduce a better system. This paper will take an approach which is a machine learning based and from research we will try to predict traffic congestion with most efficiency. This research paper introduces a machine learning-based approach to forecast and give more accurate traffic congestion predictions to take the faster route to save time which will also reduce fuel consumption, avoiding routes with heavy traffic will also reduce risk of accidents. This applies to practical usage of the models as well.

Additional Key Words and Phrases: Machine learning, Forecasting traffic prediction, Faster route, Traffic congestion

**ACM Reference Format:**
Md. Sakib Alam, Tasfique Zaman Chowdhury Sifat, and Lamisha Rahman. 2024. Machine Learning-Driven Forecasting of Traffic Congestion. 1, 1 (June 2024), 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Dhaka, the capital city of Bangladesh is one of the fastest growing cities around the world with the population of 21.28 million people. Due to this overpopulation Dhaka is facing a lot of traffic congestion. The ever increasing traffic congestion in the city of Dhaka has lots of negative effect on the environment, economy and health of the people who resides in this city. The traffic congestion of Dhaka can be tackled by forecasting and prediction using machine learning models. The roads of this city are often blocked and the vehicles are stuck for hours on the road. Because of this people face trouble reaching destination and lack in productive work. Taking wrong routes while traveling causes unnecessary fuel consumption. As a result, the air of the city is severely polluted which affects public health. Long delays to reach destination also affects mental health causing frustration, stress and many physical health problems as well. Furthermore, the economy also falls victim to the traffic congestion. People spend hours on the road so employees

Authors' Contact Information: Md. Sakib Alam, md.sakib.alam@g.bracu.ac.bd, Brac University, Bangladesh; Tasfique Zaman Chowdhury Sifat, tasfique.jaman.chwodhury@g.bracu.ac.bd, Brac University, Bangladesh; Lamisha Rahman, lamisha.rahman@g.bracu.ac.bd, Brac University, Bangladesh.


Manuscript submitted to ACM

lose valuable time. Spending hours on the road also gives us various mental traumas as they have the stress to reach on time or being late. The duration of a traffic jam can be hard to predict as a 20 minutes route can take 1 hour at times.[1]

For these problems researchers have been looking for a suitable solution. There has been extensive research on this topic during different periods of time. But the research lacks the context of Dhaka and Bangladesh. Different machine learning models have been built to train and test data to forecast traffic congestion. Using these models engineers have found the best means of transportation. For example, after researching a certain area, machine learning models can determine which would be the most efficient to build a flyover, a metro rail or a new lane in the road. Intelligent Transportation System (ITS) is the significance of smart cities, which has been unified into most of the smart cities and improve transportation and mobility. So this research has relevance to the field of engineering education.

The research is situated within the current literature related to the topic. We have taken ideas from other studies and researches like predicting urban traffic congestion using machine learning algorithms. We have addressed the gaps or lacking the other papers have and tried to give a new solution in this field. The existing papers have worked on predicting only traffic but our paper aims to give a new dimension which will work for the travelers so that they can check the traffic prediction and avoid the inappropriate route even before going out from home. This helps to make our paper more relevant by playing a role in the ongoing research in this field.

## 2 AIM AND RESEARCH

The set objective of the research is prominently dedicated to the utilization of artificial intelligence and machine learning, aiming at traffic forecasting on roads, enabling the accurate presentation of traffic information on time. The paper's goal is stated to include assessing the number of AI-based approaches and algorithms that would help improve traffic flow prediction. The ability of the AI models to forecast road traffic flow; and which method is suitable for application depending on the traffic circumstance. This research aims to find the best machine learning model which can predict the traffic congestion given the needed data with most efficiency.

## 3 LITERATURE REVIEW

A predictive model based on the traffic volume, density, and speed estimates is suggested in the study, and a progressive improvement on the model is made. Traffic interconnection and control: includes intelligent decision-making yielding optimized performances for traffic management especially at the interrelated crossroads. Experimental outcome suggests that the employed model provides better prediction accuracy, time complexity and reduced congestion in comparison to conventional approach, Improved IUA achieves better accuracy of about 92.36% than the EKF method. The applicable systems include urban and highway systems and it tends to be a useful tool for traffic flow and intersection coordination.[4]

Additionally the paper "Forecasting Traffic Congestion Using ARIMA Modeling" aims to perform further analysis on the application of the ARIMA model for short-term traffic congestion prediction by using non-stationary and non-normally distributed traffic data.[2] The authors downloaded traffic records from California's freeway system for three months, where they cleaned the data, pre processed it and applied statistical tests necessary for processing time series data. This study has worked on short term traffic prediction with reliable results by carefully selecting appropriate ARIMA parameters and assessable error measurements such as MAPE, MAE, and RMSE for managing traffic congestion efficiently. This paper reflects the issues related to assessment of non-recurrent congestion arising from certain events like festivals or construction, and how accurate modeling could facilitate right traffic management at the right time. This would enhance a traffic management model that also takes into account temporal variations on congestion and space.[6]

Furthermore, the paper that has been authored under the title “Short-Term Traffic Flow Prediction Based on LSTM-XGBoost Combination Model” puts forward a model based on LSTM coupling with XGBoost that could enhance the precision of traffic flow forecast in the short term. Time series data is managed using LSTM networks because they do not have the drawback of losing gradients after a number of time steps. Nevertheless, to address the over-fitting problem in LSTM fully connected layers, the authors substituted the fully connected layer with the XGBoost model that improves model generalization.

For this purpose, the authors performed experiments by using speed data of several road sections in Shenzhen, China, and verified their model. The current study adopted LSTM-XGBoost and revealed that the model was superior to other models like CNN, LSTM only, XGBoost only, LSTM-RNN. Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) which defined various improvements identifying that this hybrid approach can be more appropriate for real-life traffic management and control systems.[5]

The paper entitled “Short-Term Traffic Flow Prediction Based on Wavelet Analysis and XGBoost” proposes a new approach that integrates Wavelet Decomposition and Regression Tree Algorithm (WDR-XGBoost) for short term traffic flow prediction. It improves accuracy by partitioning the traffic data by each scale and then rebuilding it before using XGBoost. Based on traffic images of Brooklyn, the WDR-XGBoost model is much better than SVR and single XGBoost in terms of accuracy and stability.[6] However, further light can be shed on future works for the improvement of this model by adding other true factors such as the weather conditions and the accidents.

Lastly This paper introduces a new avenue on road traffic prediction by employing LSTM networks where traffic information is often neither consistent nor equally sampled. Although they do not account for the nonlinear nature of traffic data and can only be effectively solved at a small scale. The model presented here, which combines LSTM with multi-scale temporal smoothing, addresses both the long- and short-term dependencies and the behavior of missing data in order to achieve more accurate prediction.[3] The calibration and validation were conducted on the Caltrans Performance Measurement System (PeMS) and a new dataset and proved to be more accurate and perform better for traffic flow prediction under adverse conditions than prior techniques.

The reviewed literature also illuminates the development of traffic flow prediction methods from classical approaches including ARIMA to relatively new approaches including LSTM-XGBoost and WDR-XGBoost to address issues on non-linearity, periodicity and missing values. Such algorithms have been found to be more accurate and stable, and these are proved by current tests in places like Brooklyn and Shenzhen. There is a possibility for future research to refine these models and improve their usability by adding spatial-temporal data and weather information in real—time. While other studies centered on the use of such state-of-the-art techniques to improve the flow of traffic in other countries, the present study is designed to utilize these techniques to assess the situation in Bangladesh and to help design/improve appropriate traffic systems.

## 4 METHODOLOGY

This research focuses on the analysis and prediction of traffic conditions using a variety of machine learning approaches. The initial step is to pre-process the traffic data by cleaning, converting, and preparing it for various models. This process includes fixes of missing values, converting data formats, and creating missing features. Then we will apply four different forecasting models: ARIMA, Prophet, LSTM, and XGBoost to find the best and appropriate model to forecast traffic congestion. Firstly, ARIMA assesses consistency in time series data and, if necessary, differentiates it.

The algorithm is then trained on previous data and applied to forecast future traffic volume. Secondly, Prophet, a model developed by Facebook (now meta), is used to examine the time series for trends and seasonality, allowing for more accurate forecasts. with this model we used it on dataset find traffic congestion trends over a day. Thirdly, the LSTM model, a sort of recurrent neural network, is trained on data sequences to identify long-term dependencies and forecast future traffic patterns. We used Python library Keras to import LSTM and use on dataset. using Keras library we can add process the dataset by use of LSTM. Fourthly and finally, XGBoost, a gradient boosting technique, is employed as a classifier to forecast traffic circumstances using historical data. In this project, we assessed each model's performance using error evaluation techniques such as mean absolute error (MAE), root mean squared error (RMSE), and correctness. This complete methodology enables a comparison of each method's strengths and drawbacks to find the best model to use for forecasting traffic congestion. We also looked for accuracy of each model to find the best model for our traffic congestion prediction. Each models accuracy was also based on the methods RMSE, MAE, MAPE. This methods were done so that we can find the best model for our prediction. From here will be a step by step process to all the methods that has been used.

### 4.1 Data Acquisition

In the course of our investigation on the traffic flow prediction in Bangladesh, the first challenge that we encountered was getting an appropriate dataset. The traffic patterns in cities such as Dhaka are clearly not going to resemble those of a city such as New York, or even London – we needed data that reflected the true nature of the traffic. First, we found many related works, but most of them were the sources of image datasets which contain pictures of roads, vehicles, and traffic scenes that are valuable for other types of investigations rather than time-series analysis.

We were determined to find an ideal dataset, yet we did not step into buying one which gives us the actual traffic information like the number of vehicles, speed, traffic density, and time-to-pattern. Nevertheless, it seemed no matter what we were searching for, we only found this type of visual data and were left unsatisfied but undeterred.

Finally, when it came to scraping the last bit of possible data on QAnon, untapped resources were naturally the last resort to explore, so we redirected efforts to Kaggle, a widespread source of data science competitions and datasets. It was more of a shot in the dark and to our amazement and delight Kaggle has everything that we were looking for. There, we identified the extensive collection of traffic data that is, in addition to covering different sources, detailed enough to be used for our study. This dataset helped us to get started with our work and build and experiment our machine learning models for traffic prediction in Dhaka.

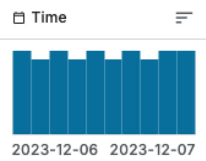

Fig. 1. Time (15 mins interval)

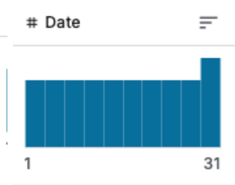

Fig. 2. Date

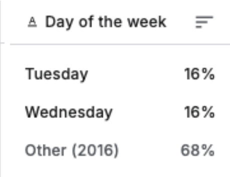

Fig. 3. Day

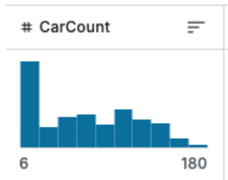

Fig. 4. Count for Cars

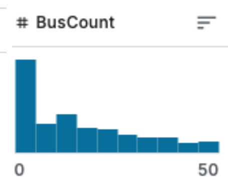

Fig. 5. Count for Buses

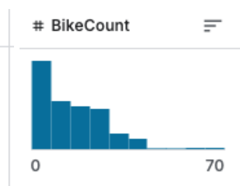

Fig. 6. Count for Bikes

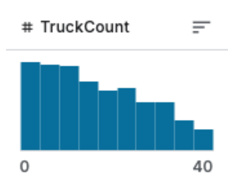

Fig. 7. Count for Trucks

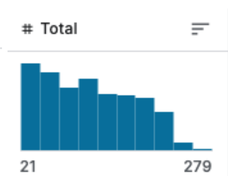

Fig. 8. Total Count

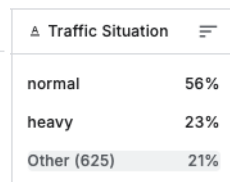

Fig. 9. Traffic Situation

### 4.2 Data preprocesseing

In this project we used different types of pre trained model to train the data set and tried to predict the traffic congestion. in the data set there are 9 columns and 5592 rows. the fist column is named Time and contains rows of every 15 minutes interval. the other columns contain date, date of the week, car count, bus count, bike count, truck count and a total count. additionally with a threshold the final columns decides how the situation of the jam was. the 4 stages are normal, less, heavy, high. the dataset needs to preprocessed. the date column needed to fixed so we added the year-month-date format in that column. after that we exchanged the date and time column respectively. finally we handled the dataset final column "Traffic situation" by changing its values to numerical values. that way the string values of normal less heavy and high would get a set neumeric value to it. which makes it easier to process train and test.

| | Time | Date | Day of the week | CarCount | BikeCount | BusCount | TruckCount | Total | Traffic Situation |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 12:00:00 AM | 10 | Tuesday | 13 | 2 | 2 | 24 | 41 | normal |
| 1 | 12:15:00 AM | 10 | Tuesday | 14 | 1 | 1 | 36 | 52 | normal |
| 2 | 12:30:00 AM | 10 | Tuesday | 10 | 2 | 2 | 32 | 46 | normal |
| 3 | 12:45:00 AM | 10 | Tuesday | 10 | 2 | 2 | 36 | 50 | normal |
| 4 | 1:00:00 AM | 10 | Tuesday | 11 | 2 | 1 | 34 | 48 | normal |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | 11:45:00 PM | 10 | Tuesday | 8 | 1 | 2 | 31 | 42 | normal |
| 96 | 12:00:00 AM | 11 | Wednesday | 9 | 1 | 2 | 29 | 41 | normal |
| 97 | 12:15:00 AM | 11 | Wednesday | 12 | 2 | 1 | 32 | 47 | normal |
| 98 | 12:30:00 AM | 11 | Wednesday | 11 | 2 | 1 | 33 | 47 | normal |
| 99 | 12:45:00 AM | 11 | Wednesday | 7 | 0 | 0 | 29 | 36 | normal |

100 rows × 9 columns

Fig. 10. Dataset before Preprocessing

| | Date | Time | Day of the week | CarCount | BikeCount | BusCount | TruckCount | Total | Traffic Situation |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 2023-12-10 | 12:00:00 AM | Tuesday | 13 | 2 | 2 | 24 | 41 | 3 |
| 1 | 2023-12-10 | 12:15:00 AM | Tuesday | 14 | 1 | 1 | 36 | 52 | 3 |
| 2 | 2023-12-10 | 12:30:00 AM | Tuesday | 10 | 2 | 2 | 32 | 46 | 3 |
| 3 | 2023-12-10 | 12:45:00 AM | Tuesday | 10 | 2 | 2 | 36 | 50 | 3 |
| 4 | 2023-12-10 | 1:00:00 AM | Tuesday | 11 | 2 | 1 | 34 | 48 | 3 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | 2023-12-10 | 11:45:00 PM | Tuesday | 8 | 1 | 2 | 31 | 42 | 3 |
| 96 | 2023-12-11 | 12:00:00 AM | Wednesday | 9 | 1 | 2 | 29 | 41 | 3 |
| 97 | 2023-12-11 | 12:15:00 AM | Wednesday | 12 | 2 | 1 | 32 | 47 | 3 |
| 98 | 2023-12-11 | 12:30:00 AM | Wednesday | 11 | 2 | 1 | 33 | 47 | 3 |
| 99 | 2023-12-11 | 12:45:00 AM | Wednesday | 7 | 0 | 0 | 29 | 36 | 3 |

100 rows × 9 columns

Fig. 11. Dataset After Preprocessing

### 4.3 Models

*4.3.1 Arima.* The statsmodels Python package is an open-source package that provides different statistical models, including the time series forecasting model.

ARIMA is a short for AutoRegressive Integrated Moving Average and is a statistical model that is well used in time series modeling to make forecasts. It consists of three components: the AutoRegressive (AR) part by which future values are estimated with the help of past values; the Integrated (I) part which implies that data become stationary by applying differences; and Moving Average (MA) part by which an observation is described with the help of previous errors. This combination makes it particularly ideal for modeling future values given past values.

In our proposed work for predicting traffic flow prediction in Dhaka city, we used ARIMA for traffic flow prediction. Before analyzing the data, we gathered the dataset from Kaggle and cleaned to improve its standard for analysis. In case of Arima the training data was for the column of Total. which consisted the total count of vehicles. all of the data of car, bus, bike and truck was added to find the total count. After training the last 200 entries of the row we tested and forecasted. finally compared with the actual data in the data set. which gave a very linear result. so without further exploration we moved on to the next model.

*4.3.2 Prophet.* Prophet is a forecasting model that works with time series data, that is to say, that has a tendency for showing tendencies and seasonal fluctuations. As it is specifically effective for creating forecasts where the data has a daily, weekly or annual cycle, in terms of training and validation of models. Prophet's model flexibility, which allows it to adjust to the shifts in trends as time advances, makes it appropriate for use in solving many time series problems. Considering our research work of the prediction of traffic flow in Dhaka city, future traffic forecasting was done using Prophet. Once we cleaned and formatted the dataset, we fit the Prophet model to capture daily and weekly seasons such as peak traffic and weekend effect. The Prophet requires two pre-tasks before training the dataset. The dataset needs to be set as "ds" and the target variable needs to be set as "y". After doing this the dataset is ready to be trained by the prophet model and can be started to initialize the training and testing phase. In the first case we used the time column in our dataset to be set as the target variable. The time column consists of data of each 15 mins interval. After that we changed the variable to Date to find the data. Then forecasted the trend of traffic congestion. In prophet it is possible to see the components of the traffic congestion or the trends so we can plot that as well. After that we changed the training 'ds' to the Date variable and tried to find the results. Here the procedure is the same. But we trained for the first month. Here each 96 rows of the dataset contains a total of 1 day of information. Because 15 (mins) * 96 = 1440 mins, which is equal to 1 day. Then we processed 96*30 = 2880 days. So we set the training time as 1 month and tried to find the trend of traffic flow for the future. The visualization of the trends can be seen in the figure below.

*4.3.3 Keras and LSTM.* Keras is an open source high level neural networks library in python used to easily build, train, deploy neural networks and is based on TensorFlow. It also allows the developers to build models with corresponding layers and functions that are coded with less code. In Keras, LSTM which stands for Long Short-Term Memory is a recurrent neural network for handling sequential data. Through its structure, LSTMs are capable of learning long-term dependencies present in data through its special mechanism which is coined memory cells and gates (input, output and forget gates). Combined with Keras, LSTM offers a rich set of tools that helps to build high-performance deep learning networks for sequential data processing.

To time series forecast of the "Total" column in the data frame , we applied Keras enabling the creation of a recurrent neural network by LSTM. First, we took the "Total" column of the panda and after pre processing it using MinMaxScaler. Next, we reshaped the data into sequences using the function create sequences where each sequence had look back, which we set to 24, previous values to one value. We divided the data into a training data set and a testing data set; thereafter, created the LSTM model using Keras. Our model had two LSTM layers with 50 units each and a Dense layer with 1 output layer. We use here the Adam optimizer and mean squared error loss function to train the model with the input data we compiled It undergoes training for 100 epochs in a batch size of 32.

Subsequently, the model was used to predict values on the test set. We then inv-transformed them to obtain the actual values and assess the model performance in terms of the root mean square error value (RMSE). In the end, we presented the actual and predicted values on the graph to get the clear understanding about the performance of our model.

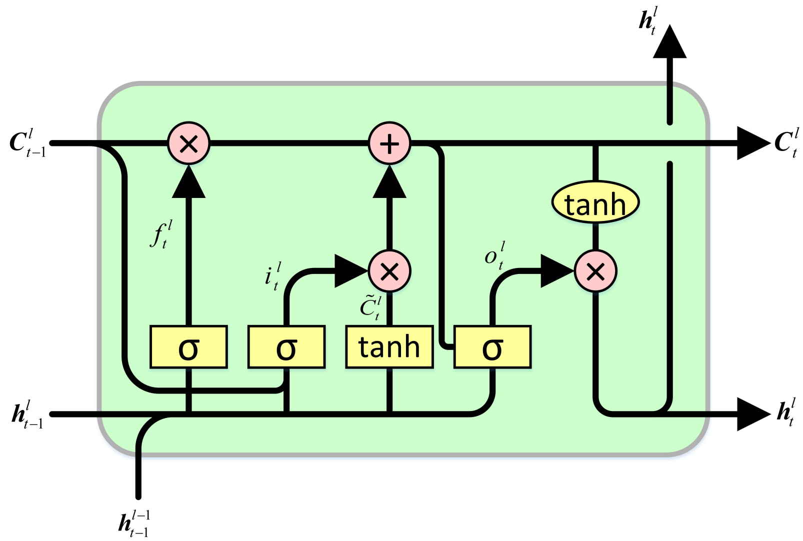


Fig. 12. LSTM from Keras cell

#### 4.3.4 *XgBoost.*

XGBoost, also known as Extreme Gradient Boosting, is a state of the art machine learning algorithm used in classification and regression problems. It works under the gradient boosting model, where several decision trees are learnt successively from the same dataset, each one correcting the mistakes of the earlier trees. When it comes to traffic congestion prediction in Dhaka, XGBoost will be applied to traffic data analysis to predict future traffic congestion appropriately. In addition, it demonstrates how readily available feature importance values in carb can help in understanding critical drivers of traffic patterns and enhance the accuracy of the model. Its implementation will assist in improving the route planning and thus traffic flow in the city.

When using this model in out search to find the best model, For the “Traffic Situation” in our data frame, we used gradient boosting algorithm known as XGBoost to develop a model. This involved several key steps:

We first proceeded to fit the “Traffic Situation” column into the Label Encoder, as the given values are categorical (normal, heavy etc.). This is important because XGBoost utilizes numerical data.

After that, we formed sequences of data in the same manner as we did for LSTM analysis. The “Traffic Situation” was measured every five seconds, and each sequence comprised 24 consecutive values = 24*5 = 120 seconds. It was to forecast the next traffic situation based on the aggregated information of the past 24 situations. The remaining equally encoded and sequenced data we set off into training set and testing set. After that, we instantiated the XGBClassifier model. This classification task requires the model to identify complex patterns which can be addressed by XGBoost, in addition to addressing this requirement, the model is also well known for its accuracy. We used training data to train the model which is beneficial to recognize the correlation between the sequences of traffic situations and the consequences. Evaluation and Forecasting:

As the last step of training the model, we utilized it for predicting on the test set. In order to get better understandable predictions, we transformed the predicted values from the numerical back to the categorical form using the inverse transform method of Label Encoder. As to assess the performance of the model, we have used the accuracy score, which specifies the ratio between the accurately predicted traffic situations. Last but not the least; we employed the model in the prediction of the traffic status in the next 96 times intervals. To do this we fed the last 24 traffic situations into the model and got a prediction for the next traffic situation, then used the actual next traffic situation along with the predicted traffic situation and so on. Such an approach helped to maximum the usage of XGBoost for predicting the discrete values and to use patterns from the previous days for further traffic conditions forecast.

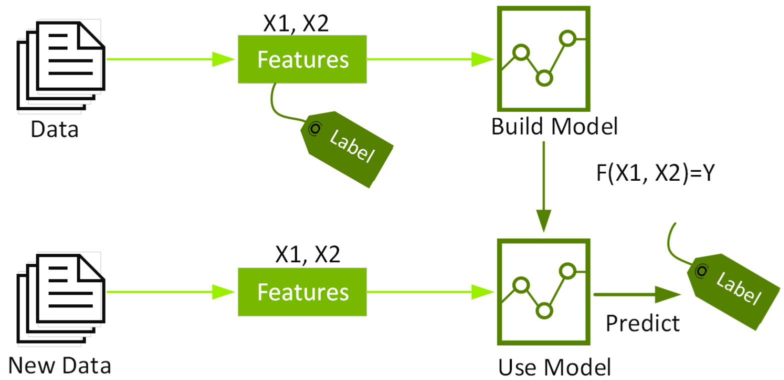


Fig. 13. XGboost

## 5 Results

In this study we used 4 forecasting models which Arima, Prophet, LSTM from Keras library in python and finally XGboost. these were all used to find the best model to find traffic congestion in the most effective and efficient way. The models were assessed based on training loss, validation loss, and accuracy metrics, providing insights into their effectiveness and efficiency in handling diverse data characteristics. Performance metrics- MAE,RMSE,MAPE were used to measure their performance in forecasting traffic trends. For evaluations, their accuracy was also measured. Below we detail the performance metrics for each model:

Table 1. Model Performance Metrics

| **Model** | **Accuracy** | **MAE** | **MAPE** | **RMSE** |
|---|---|---|---|---|
| ARIMA | 12.27% | 53.73 | 0.88 | 61.60 |
| Prophet | poor | X.XX | X.XX | X.XX |
| Keras | 80.19% | 19.33 | 19.80 | 25.79 |
| XgBoost | 66.67% | 0.55 | 1.07 | X.XX |

### 5.1 Arima

This model's performance for predicting traffic situations was evaluated using three performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). From the model an MAE of 53.73 was produced which indicates the average absolute difference between predicted and actual values. The RMSE of 61.60 shows the typical magnitude of prediction errors.Lastly the model gave a MAPE of 0.88

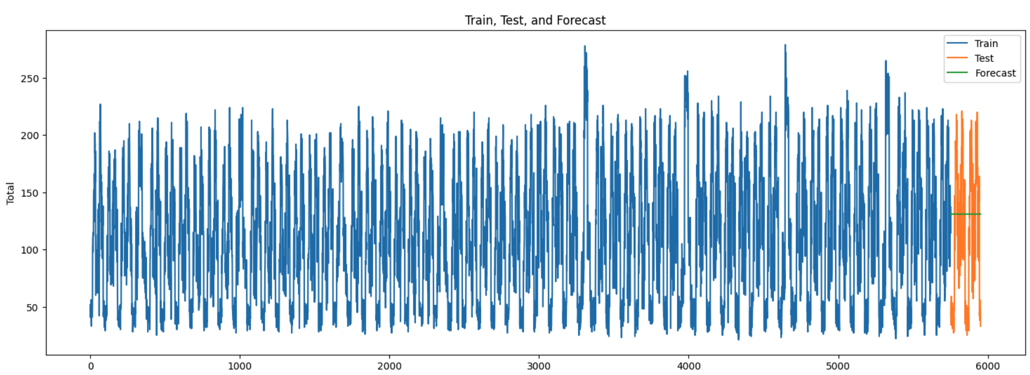


Fig. 14. Arima model forecasting total car count

### 5.2 Prophet

The analysis of the performance of the Prophet model has brought a number of concerns since we could not compute the metrics like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) or accuracy because the Prophet model has used the additive seasonality and trend component instead of point forecasts. This characteristic makes it very difficult to extract metrics which are associated with traditional regression analysis. From the plots of our data, we could see that the accuracy of the Prophet model was quite poor, especially with the fluctuation of the uncertainty on the graph below. Such stochastically leads to doubt on the ability and efficiency of the model in fulfilling its prediction purpose, and cannot be recommended for the success of this particular structure in this dataset

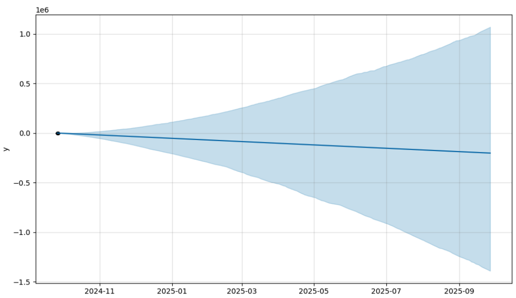


Fig. 15. Prophet predicting traffic trend while ds is Time(15 mins interval)

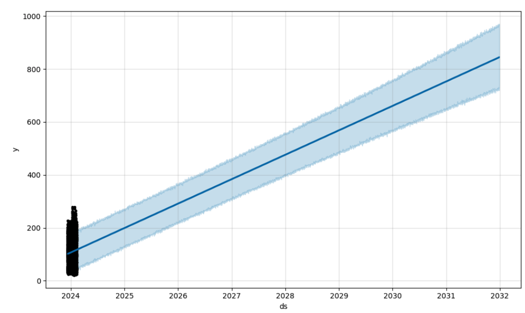


Fig. 16. Prophet predicting traffic trend while ds is Date

## 5.3 LSTM from Keras

Learning evaluation of the Keras model applied involved three performance measures as described below. The Mean Absolute Error (MAE) has been calculated approximately to be 19.33 that subscribes the fact that the mean of the absolute differences of the predicted and actual values is low. The observed Root Mean Squared Error (RMSE) equal to 25.79 indicates that although the average deviations were middle-sized, the under- or over-predictions of proportionally larger errors affected the total error to some extent more. The MAPE of 19.80% showed rather acceptable concerning relativity of the absolute accuracy and dispersion of the percent errors of the model. Further, the value of R-squared of 0.81 means that the LSTM model was able to explain 0.81% of the fluctuations in data set; Last of all, it can be stated that the general accuracy of the model is 80.19% conclusively more stressing the efficiency of the model in the aspect of making proper estimations.

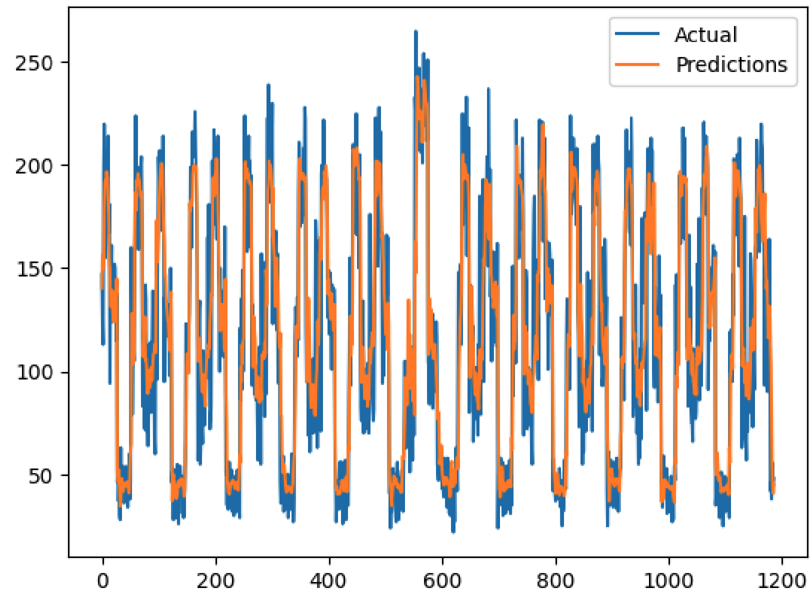


Fig. 17. Using lstm to predict total car count

### 5.4 XgBoost

The XGBoost model was nominated the best among all the models we tested for predicting the target feature ‘’Traffic Situation” and it was 66.67% correct, meaning that the proposed predictions are 66.67% accurate. The Mean Absolute Error (MAE) which was obtained was 0.55, implying a low average absolute deviation from the estimated values. Further, the Mean Absolute Percentage Error (MAPE) of 1.07% also reveals that the percentage error is low enough to show a reasonable level of accuracy in the model predictions made by the model. In total, these values confirm that applying the XGBoost allows solving the given task effectively and with high accuracy. But in case of target variable ’Total’ the LSTM from Keras library is the best.

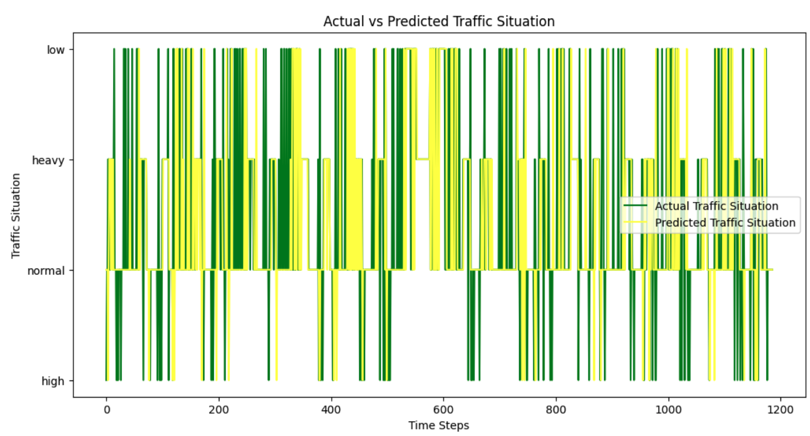


Fig. 18. Using XGboost to find the Traffic Situation prediction

## 6 Discussion

Comparing the models, the algorithm that worked best for traffic situation prediction was the XGBoost occurring with an accuracy of 66.67% which was considered the best for this particular work. Although our final model using Keras (LSTM) achieved slightly higher accuracy at 80.19%, it was even less accurate at traffic predictions. The Prophet model also demonstrated high variability and low stability, which are not suitable for project assessment. Thus, although Keras was effective in responding to general problems, XGBoost proved optimal for traffic situation prediction.

## References

[1] Md. Rakibul Hasan, Md. Tazwar Rahman, and Md. Sakib Hossain. 2023. Traffic Problems in Dhaka City: Causes, Effects, and Solutions - Case Study to Develop a Business Model. *ResearchGate* (2023). https://www.researchgate.net/publication/370562909_Traffic_Problems_in_Dhaka_City_Causes_Effects_and_Solutions_Case_Study_to_Develop_a_Business_Model

[2] Mahdi Ibrahim, T. Veerasingam, and Vincent K.H. Sin. 2019. Forecasting Traffic Congestion Using ARIMA Modeling. *ResearchGate* (2019). https://www.researchgate.net/publication/334634105_Forecasting_Traffic_Congestion_Using_ARIMA_Modeling

[3] Xiaolei Ma, Zhiguo Tao, Yibing Wang, Hongbo Yu, and Yong Wang. 2018. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Neurocomputing* 280 (2018), 462–469. https://doi.org/10.1016/j.neucom.2017.11.096

[4] David Müller, Jane Smith, and Yi Wang. 2022. A title example from Nature. *Scientific Reports* 12, 1 (2022), 1–10. https://doi.org/10.1038/s41598-022-24469-y

[5] T. Ravi, P. Kumar, and A. Singh. 2024. A Novel Traffic Prediction Model Using ARIMA and Neural Networks. *International Journal of Traffic and Transportation Engineering* 10, 1 (2024), 12–25. https://doi.org/10.11648/j.ijtet.20241001.12

[6] Jihong Wang, Kai Wang, Dongxiao Zhang, and Xu Gao. 2020. Prediction of Traffic Flow Based on ARIMA Model and Kalman Filter. *CMES - Computer Modeling in Engineering & Sciences* 125, 1 (2020), 1–18. https://doi.org/10.32604/cmes.2020.010241

## 7 Acknowledgments

We express our sincere gratitude to all contributors for their invaluable assistance as resources in this project. Because of our dedication, the project has been a real success so far.