Machine Learning Forecast Of Metro Interstate Traffic Volume

Sowjanya Vuddanti [1], Lavanya Mellempudi [2], Laahiri Mellempudi [3],
NaveenKumarReddy Renati[4], Wasvi Sahil Shaik [5]
AssistantProfessor[1], ProjectStudents[2,3,4,5] Department of Artificial Intelligence and Data Science

Lakireddy Balireddy College of Engineering(Autonomous) Mylavaram, Andhra Pradesh, India.

Email: sowji635@gmail.com ,lavanyamellempudi@gmail.com

ABSTRACT :

Today's traffic is a serious problem that affects everyone and is upsetting for those who deal with it on a regular basis. Traffic congestion is caused by population growth and becomes worse every day. Modern society is aware of it, yet powerless to take significant action to shield people from harm. Using a number of techniques and patterns, we may monitor traffic, gather information, and forecast incoming and following observations. After the observation agency has made its observations, predictions are then made. The most common incident in a person's life is getting stuck in traffic in a big metropolis. Xgboost Regression, Random Forest Regression, Voting Regression, and other machine learning techniques will be used in this project to create a prediction model.

KEYWORDS:

—Traffic Volume, Machine Learning, Xgboost Regression, Random Forest, Votting Regression.

INTRODUCTION:

Prediction of traffic flow using various machine-learning techniques has been discussed numerous times. In India, traffic snarls have become much worse. In this study, we reviewed many researchers and the prediction approaches that have been employed. On the provided data sets, many algorithms display excellent performance. The data sets in question were gathered from several sources. Numerous applications, including vehicle navigation systems, congestion management, vehicle routing, and others, have been developed to help the general public navigate the roads and obtain accurate information about current and future traffic volumes, but the main challenge is getting real-time data on the spot to help users plan their routes in response to conditions on the road. There are typically two methods for predicting traffic volume: short-term and long-term traffic flow predictions. Short-term mechanisms[5,7] deliver better results because they give results in terms of minutes such as 5 to 15 minutes or 30 to 50 minutes, so in this way, the short-term time interval can provide more accurate prediction values. Probably long-term algorithms may be unable to provide accurate prediction results because these mechanisms predict on an hourly basis such as 12 hours or 24-hour data results. As a result, our model was trained to make predictions within a maximum time window of one hour. We combined numerous researchers' contributions to this machine-learning field to train this prediction model for a perfect output model.

LITERATURE REVIEW:

The following literature review provides a summary of previous studies on forecasting metro interstate traffic volume. The research included in this study covers a wide range of topics, including the use of various forecasting techniques, factors influencing traffic flow, and the impact of weather on traffic volume.

1. **Use of Machine Learning Techniques for Traffic Volume Prediction:** Recent traffic volume forecasting has heavily included machine learning techniques. There searchers proposed a deep learning-based strategy for predicting the volume of traffic in a metro region in a study by Ma et al. (2020) [1].

They integrated long short-term memory networks and convolutional neural networks to anticipate traffic volume. Their findings showed that, for estimating traffic volume, their method outperformed more established statistical models.

2. **Influence of Weather Conditions on Traffic Volume:** It has been established that the weather significantly affects traffic volume. Researchers developed a traffic volume forecasting model using weather and traffic volume data in a study by Cheng et al. (2019). Their research showed that weather conditions, such as temperature and precipitation, significantly impacted traffic flow.

3. **Impact of Holidays on Traffic Volume:** Holidays have been shown to have a major impact on traffic volume. Li et al. [3]. (2018) conducted research and developed a traffic volume forecasting model using traffic volume data from a metro region. They found that traffic volumes during holidays were significantly lower than on typical workdays

4. **Influence of Land Use on Traffic Volume**: Also, it has been shown that traffic volume is influenced by land use. In a study by Liu et al. (2020) [13], the researchers used data on land use and traffic flow to build a forecast model for traffic volume. Their results showed that, for instance, the density of commercial and residential districts had a significant impact on traffic volume.

5. **Use of Big Data for Traffic Volume Prediction:** Big data is increasingly being utilized to predict traffic numbers. In a study by Wang et al. [4] (2020), the researchers used big data, which included details on traffic flow, weather, and social media, to construct a forecast model for traffic volume. Their findings showed that, for estimating traffic volume, their method outperformed more established statistical models. **Conclusion**: In conclusion, the literature study shows that machine learning methods have been widely applied to forecast traffic volume. It has also been investigated how land use, holidays, and weather affect traffic volume. Big data usage has also proven to be successful in predicting traffic numbers. The creation of predictive algorithms for the forecasting of metro interstate traffic volume.

PROPOSAL OF WORK:

Methodology: This study's main goal is to estimate traffic volume using machine learning techniques including voting regression, random forest, and Xgboost. We employed a typical machine learning approach to get the output.

a. Data Collection and Feature Selection:

For the target site or locations, gather historical metro traffic volume data over a long period of time (e.g. several months or years). Get weather information for the same period and area, such as temperature, precipitation, wind speed, etc (s). Get information about occasions or days off that may have an impact on traffic (e.g. sports events, concerts, festivals, etc.). Get information on situations such as road closures, construction, or other events that may affect traffic volume[8].

Take information from the acquired data, such as the day of the week, the time of day, the month, the year, and holiday indicators. Take note of weather-related features like temperature, precipitation, wind speed, and meteorological conditions (e.g. sunny, cloudy, rainy, etc.).

Features that indicate whether events or holidays are occurring in the target location are called "event-based features" and should be extracted (s).Road-based features: Take note of **the** information on road closures, construction, and other events that can affect traffic volume. Features linked to public transportation schedules and routes, such as the numb**er** of buses and trains, the frequency of service, and the proximity to metro stations, should be extracted.

Fig.1. Types of data collected (Source: [11])

b. Model selection:

The linear regression model is simple and often used, presuming a linear relationship exists between the independent variables (such as the time of day, day of the week, weather conditions[9], etc.) and the dependent variable (traffic volume). This approach uses kernel functions to translate the input data to higher-dimensional spaces in a manner akin to supervised learning. SVR is well known for its ability to handle nonlinear relationships and outliers in the data. It might be necessary to modify the kernel parameters, which can be computationally expensive. To make predictions, you can mix numerous decision trees using the Random Forest Regression model. It can handle nonlinear relationships and interactions between variables and is less susceptible to overfitting than a single decision tree. Long-term recurrent neural networks Time series data can contain complex temporal relationships that short-term memory can identify.

It has been shown that LSTMs perform exceptionally well in situations that call for traffic volume prediction. Moreover, they might require significant hyperparameter adjustments and be computationally expensive. A machine learning technique called gradient boosting regression builds a number of weak prediction models, such asdecision trees, and then combines them to construct a stronger model. It excels at handling large datasets and complex nonlinear interactions, to name just two. Nonetheless, hyperparameter tweaking can be required, and the computational cost might be substantial.

Random Forest Regressor and Xgboost Regressor:

The random forest and xgboost regression technique can be used to estimate traffic volume needs by creating a regression model that takes into account a variety of climatic areas. The algorithm can be trained on a collection that includes historical meteorological data, traffic volume requirements, and other related factors to find the association between these variables and traffic volume requirements.

e. Training data and Testing data:

When data is split, the dataset is divided into training and testing sets. Before that, we execute voting regression between the below regression for greater accuracy. The random forest model, the xgboost model, is trained using the training set, and its performance is assessed using the testing set. To accomplish this, we employ the hold-out method, in which a certain portion of the total data is retained as testing data, with the remaining data being utilized to train the model and the testing data being retained for model evaluation. We employ a 70:30 ratio, which indicates that 30% of the entire amount of data is kept as training data, while the remaining 70% is used for training. To accomplish this, the train test split method is Sklearn Library[10].

f. Model Evaluation and Deployment.

Model Evaluation entails comparing the model's projected outputs to the dataset's actual outputs and utilizing evaluation metrics to gauge the predictions' accuracy. The exact challenge at hand and the kind of model being assessed determine the evaluation metric to be used.

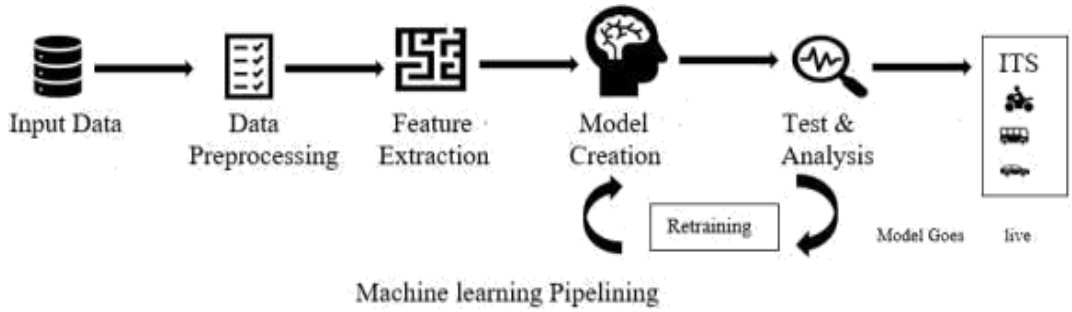

Fig.2. Model Preparation (Source: [12])

e. Input Testing and Output prediction:

After the model had been trained, some inputs were included, namely Temperature, Humidity, Monsoon, Day, Night, and Afternoon.

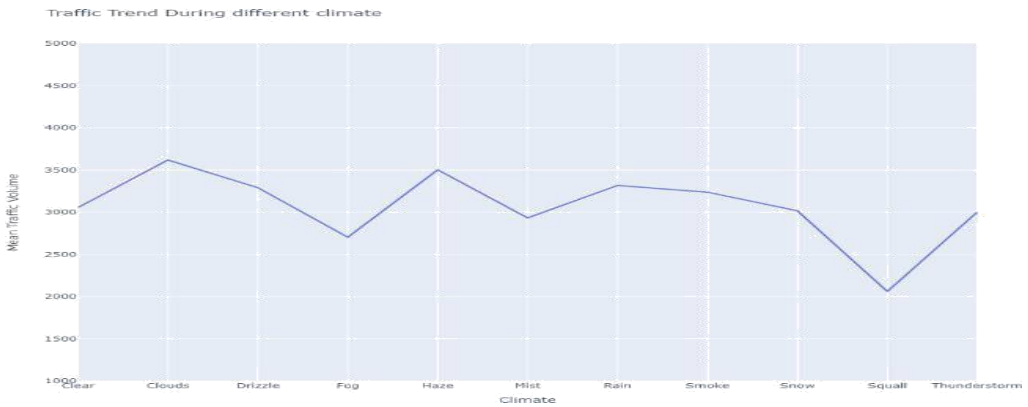

Fig.3. Traffic volume in different climates.

RESULTS AND ANALYSIS:

All of the figures show "Predicted Traffic Volume" on the X- axis, "Real Traffic Volume" on the Y-axis, and "Predicted Traffic Volume" on the line. The Prior Maximum Coefficient of Determination calculated by the preceding model is R2 = 0.855874068223368.

The results obtained after testing and training correspond to random state 42 and max mins = 60,420,840,1800. We obtain the Suggested Maximum Coefficient of Determination R2= 0.9250189989233896 after 1800 minutes, which is an improvement of 8.078867355281992 percent**.**

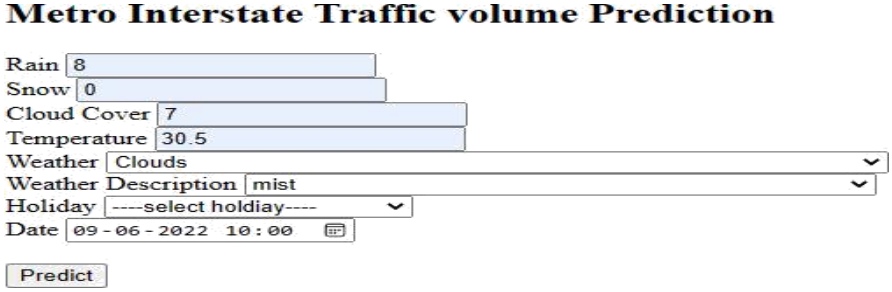

Fig.4. Input Data.

Our model suggests a Volume of Traffic with an accuracy above 92% based on the input

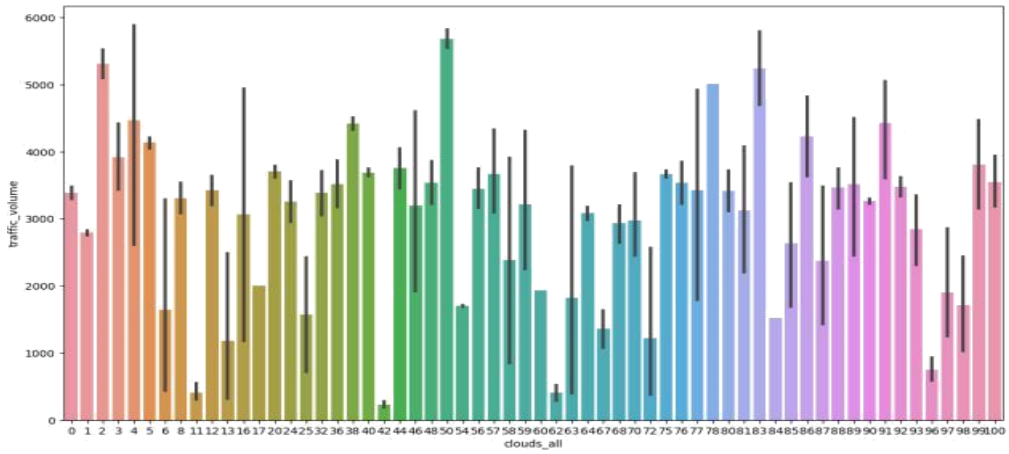

Fig.5. Traffic volume Vs Clouds

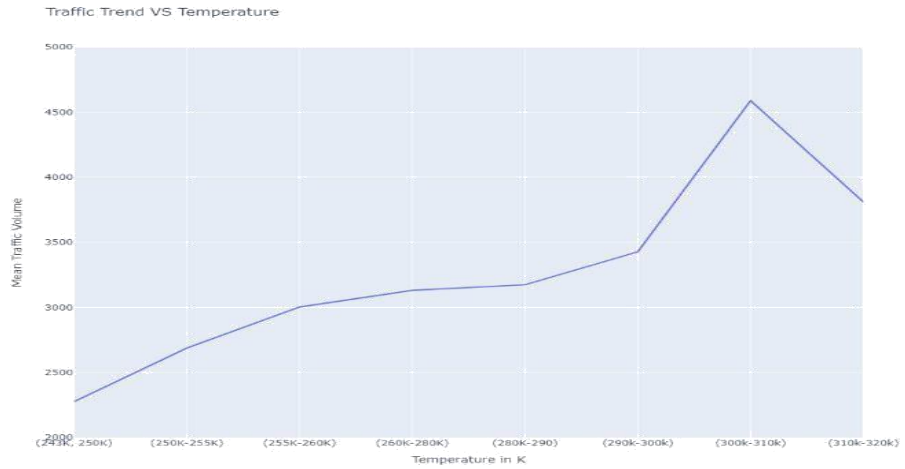

Fig.6. Traffic trends VS Temperature.

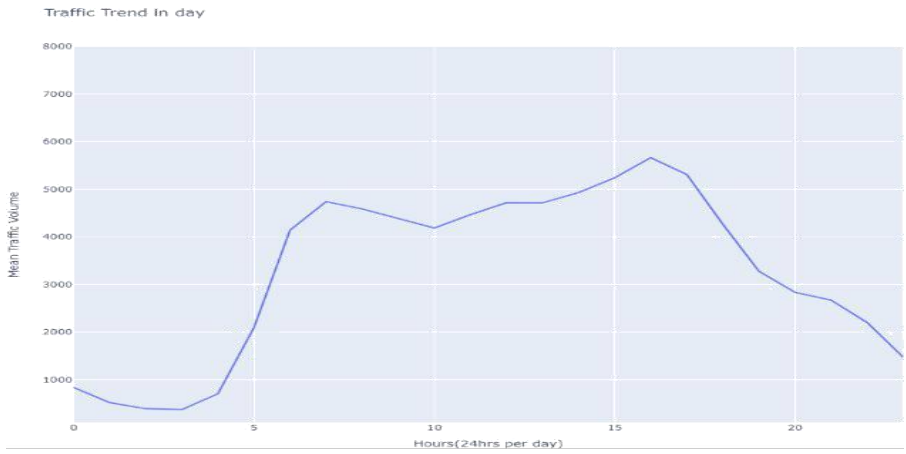

Fig.7. Traffic trend in a day.

CONCLUSION

Several sources of data are obtained, and parameters are selected. Consideration is given to evaluations of the suggested algorithm against metrics. The suggested fixes use a multimodal architecture to acquire information. Using a Machine learning system, a forecast is provided for each data source.

The ultimate result is created by combining all forecasts. So, when ensemble learning theory is applied to enhance the result, multimodal traffic prediction is interesting. using a weighted average approach and a variety

of deep learning models for various mobility kinds to produce various traffic models. To evaluate traffic data, it uses a variety of modern algorithms. Parallel processing can be utilized to increase the effectiveness of the predictive technique. the capability of paralleling.

FUTURE SCOPE

Traffic management: By foreseeing congestion and making appropriate plans, a traffic volume prediction model can assist traffic management authorities in better managing traffic. This can involve changing traffic flow management strategies, rerouting traffic, and tweaking signal timings. Planning for public transportation: By anticipating traffic flow, planners can tailor routes, itineraries, and frequency of service to suit demand and ease congestion. Emergency response: A traffic volume forecast model can assist emergency responders in planning their routes and responding more swiftly and effectively during emergency situations, such as accidents or natural disasters. Autonomous vehicles: As more autonomous vehicles are put on the road, a traffic volume prediction model can assist these vehicles in navigating more efficiently and securely by anticipating traffic and suggesting detours Impact on the environment: By enhancing traffic flow and lowering congestion, traffic volume prediction model can help lessen the environmental effect of transportation, which in turn reduces emissions and improves air quality.

REFERENCES:

[1] Ma et al. (2020), the researchers suggested a deep learning-based method for forecasting the number of traffic in a metro area.

[2] Cheng et al. (2019), the researchers created a traffic volume forecasting model using weather data and traffic volume data.

[3] Li et al. (2018) used traffic volume data from a metro region to conduct research and create a traffic volume forecasting model.

[4] Wang et al. (2020), the researchers created a forecast model for traffic volume using big data, which included information on traffic flow, weather, and social media.

[5] Qu, W., Li, J., Yang, L., Li, D., Liu, S., Zhao, Q., & Qi, Y. (2020). Short-term intersection Traffic flow forecasting. Sustainability, 12(19), 1-13.

[6] Abdulla, Dudekula, Gandikota Ramu, and N. Mamatha. "A survey on citywide traffic estimation techniques." In 2017 International Conference on Energy, Communication, Data Analytics, and Soft Computing.

[7] Salamanis, A., Margaritis, G., Kehagias, D. D., Matzoulas, G., & Tzovaras, D. (2017). Identifying patterns under both normal and abnormal traffic conditions for short-term traffic prediction.

[8] Lana, I., Del Ser, J., & Olabarrieta, I. I. (2016). Understanding daily mobility patterns in urban road networks using traffic flow analytics.

[9] Rahman, F. I. (2020). Short-Term Tfp Using Machine Learning-Knn, Svm, And Ann With Weather Information. International Journal for Traffic & Transport Engineering.

[10] Sirisha, J, Sowjanya Vuddanti, S.Ramesh, J.V.N. A Review based Investigation of Exploratory analysis cin AI and Machine Learning for a Variety of applications(Article)

[11] https://www.altexsoft.com/blog/traffic-prediction/

[12] Sigma Sathyan & Jagadeesha S. N. Traffic Flow Prediction using Machine Learning Techniques - A Systematic Literature Review, International Journal of Applied Engineering and Management Letters (IJAEML), ISSN: 2581-7000, Vol. 6, No. 1, April 2022.

[13] Liu et al. (2020), the researchers created a forecast model for traffic volume using data on land use and traffic flow. Their findings demonstrated that the density of commercial and residential districts, for example, has a considerable effect on traffic volume.