International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 157-162, April-2024

https://doi.org/10.47001/IRJIET/2024.804021

The Traffic Congestion Prediction Using Machine Learning

[1]*Pranit Jadhav, [2]Om Mohite, [3]Sagar Gite, [4]Dr. Sudhir B. Lande

[1,2,3,4]Electronics and Telecommunication Department, VPKBIET, Baramati, Pune, 413133, India

Authors E-mail: [1]*pranit.jadhav.entc.2020@vpkbiet.org, [2]om.mohite.entc.2020@vpkbiet.org, [3]sagar.gite.entc.2020@vpkbiet.org, [4]sudhir.lande@vpkbiet.org

***Abstract* - Around the world, traffic congestion is a major issue that affects everyday commutes, economic activity, and the sustainability of the environment. The goal of this research project is to use machine learning approaches to address the problem of traffic congestion. Through the utilization of an extensive dataset that encompasses several spatiotemporal parameters, including date, time, weather, and holiday indicators, the research creates predictive models that enable precise forecasting of traffic congestion levels. In order to determine how well machine learning predicts congestion dynamics, the research uses multiple regression methods, such as MLP Regressor, Stacking Regressor, and SVR. Historical traffic volume data is used to train and assess the models, making it possible to identify the main variables impacting patterns of congestion. The findings demonstrate how weather, time of day, and holidays all have a major impact on traffic congestion. Transportation authorities and urban planners may enhance overall urban mobility, optimise infrastructure utilisation, and regulate traffic flows proactively by leveraging the predictive capabilities of machine learning models. This work advances intelligent transport systems by offering a strong framework for anticipating and controlling traffic congestion. The results highlight the significance of utilising data-driven methodologies to tackle intricate urban issues, ultimately cultivating more effective and sustainable urban settings.**

Keywords: Regression, Congestion forecasting, Recurrent Neural Network, Urban mobility, Traffic Volume Trends.

I. INTRODUCTION

The globally, traffic congestion is a major problem that poses severe obstacles to economic productivity, environmental quality, and transportation networks. Cities' road networks are under increasing strain due to rising automobile traffic and fast urbanisation, which causes delays, annoyance, and financial losses. The creation of efficient congestion management techniques and a thorough grasp of the fundamental causes impacting traffic patterns are necessary to address traffic congestion.

Recent developments in machine learning and data analytics have opened up new ways to address difficult urban issues like traffic congestion. Researchers and transportation authorities can acquire significant insights into traffic dynamics and construct prediction models to precisely estimate levels of congestion by utilising enormous volumes of data from varied sources, including traffic sensors, weather stations, and urban infrastructure.

This research project focuses on harnessing the power of machine learning to predict traffic congestion and enhance urban mobility. By analyzing a rich dataset comprising temporal, spatial, and environmental variables, including date, time of day, weather conditions, and holiday periods, the study aims to develop robust predictive models capable of forecasting congestion levels with high precision. The main goals of this research are to: (1) examine how different factors, including weather, temporal patterns, and special events, affect the dynamics of traffic congestion; and (2) create machine learning-based models that can precisely predict the levels of congestion under various scenarios. By fulfilling these goals, this study hopes to aid in the creation of proactive plans for managing traffic and the enhancement of urban transport networks.

Through a combination of data analysis, model development, and performance evaluation, this research endeavors to advance our understanding of traffic congestion and provide practical solutions for mitigating its adverse effects. By integrating machine learning techniques into traffic management systems, cities can anticipate congestion hotspots, optimize traffic flow, and improve overall urban livability and sustainability. In the following sections of this paper, we will delve into the methodology employed for data collection and preprocessing, describe the machine learning algorithms utilized for congestion prediction, present the experimental results and performance evaluation, and discuss the implications of this research for urban transportation planning and policy-making. Ultimately, this research aims to contribute to the development of smarter, more efficient, and more resilient urban transportation systems capable of meeting the evolving needs of modern cities.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 157-162, April-2024

https://doi.org/10.47001/IRJIET/2024.804021

II. LITERATURE REVIEW

Table1: Examples of methods or algorithms for identifying trends in traffic congestion

Paper Title	Published Year	Technique	Methodology Result
"Traffic Congestion Prediction Using Machine Learning"	2024	Machine Learning (MLP Regressor)	Trained MLP Regressor model with historical traffic data
"Ensemble Learning for Traffic Congestion Forecasting"	2022	Ensemble Methods (Random Forest, XGBoost)	Combined Random Forest and XGBoost models for prediction
"Hybrid Approach for Traffic Congestion Forecasting"	2021	Hybrid model (ARIMA and MLP)	Integrated ARIMA and Multilayer Perceptron for prediction
"Deep Learning Models for Traffic Congestion Prediction"	2023	Deep Learning (MLP, CNN)	Utilized MLP and CNN architectures for prediction
"Traffic Congestion Prediction Using Support Vector Machines"	2020	Support Vector Machines (SVM)	Employed SVM with historical traffic data for prediction
"Temporal Pattern Analysis for Traffic Congestion Prediction"	2019	Time-Series Analysis	Analyzed temporal patterns using statistical methods
"Traffic Congestion Prediction Using Recurrent Neural Networks"	2022	Recurrent Neural Networks (RNN)	Implemented RNNs to capture temporal dependencies in data

III. METHODOLOGY

The project's methodology, as shown in fig. 1, uses machine learning techniques to forecast traffic congestion levels based on historical data and a variety of contextual variables, including the time of day and weather. After gathering and preparing the data, we design pertinent features and choose suitable regression models. We then train, assess, and optimize these models before integrating them into a web application framework to anticipate traffic congestion in real time. The goal of this all-encompassing strategy is to offer practical insights for urban planning and traffic management.

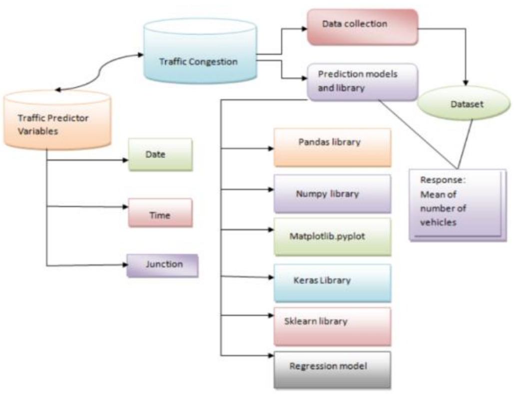

Figure 1: Traffic Congestion Prediction Methodology

3.1 Dataset and Data Pre-processing

The dataset utilized in this project on "traffic congestion prediction using machine learning" comprises 13,221 rows and encompasses various features pertinent to traffic and meteorological conditions. These features include date and time stamps, indicators for holidays, air pollution index, humidity, wind speed, wind direction, visibility, dew point, temperature, rainfall, snowfall, cloud cover, and descriptive weather attributes such as weather type and description. Additionally, the dataset contains traffic-related information such as traffic volume at the observation time and historical traffic volume for up to six previous hours. With a total of H columns, this dataset offers a comprehensive view of the factors influencing traffic congestion.

Prior to model training, extensive data preprocessing was performed, including addressing missing data, scaling features, and encoding categorical variables to guarantee alignment with the machine learning techniques used. To enable thorough model evaluation, Training and testing sets were created by further dividing the dataset. To correctly estimate traffic congestion levels, machine learning algorithms were taught utilising features including weather and historical traffic data. Throughout the study, ethical factors like as biases and data privacy were taken into mind. In order to give openness and context for the research findings, limitations

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 157-162, April-2024

https://doi.org/10.47001/IRJIET/2024.804021

inherent in the dataset, such as potential biases in data collecting or limitations in feature representation, were mentioned. All things considered, this dataset was a great tool for creating predictive models that try to improve urban mobility and traffic control.

3.2. Machine Learning Algorithms used for Prediction

The project utilizes a Multilayer Perceptron (MLP) Regressor from the sklearn.neural_network module as the primary machine learning algorithm for traffic volume prediction. The working structure of MLP-regressor model is shown in Fig. 2.Regression problems are suited for MLP regressors, a sort of feedforward neural network model that can recognise intricate patterns in data. In this project, the MLP regressor is trained on preprocessed features extracted from the dataset, including weather conditions, date and time information, and historical traffic data. By iteratively adjusting the weights and biases of the network through backpropagation, the MLP regressor learns to map the input features to the corresponding traffic volume.

Additionally, the project utilizes ensemble learning through the Stacking Regressor from the sklearn.ensemble module. The working methodology and structure of stacking regressor model is given in Fig. 3.

Ensemble learning combines multiple individual models to improve predictive performance. In this case, the Stacking Regressor combines the predictions of multiple base estimators, including the MLP Regressor, along with other regression algorithms such as Support Vector Regressor (SVR), which is imported from the sklearn.svm module. By aggregating the predictions of these base estimators, the Stacking Regressor aims to provide more accurate and robust traffic volume predictions.

The preprocessed dataset is used to train the MLP Regressor and the Stacking Regressor using features including the weather, time and date information, and historical traffic data. The system is then able to forecast traffic congestion levels depending on the features provided by using the algorithms to generate predictions on fresh input data.

The selection of these algorithms is a reflection of their capacity to identify intricate patterns in the data and offer precise traffic volume forecasts, both of which are necessary for efficient traffic congestion relief and urban mobility.

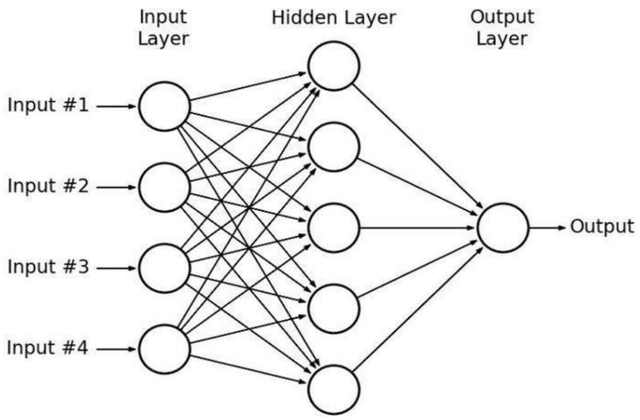

Figure 2: Multi-Layer Perceptron Regressor model structure

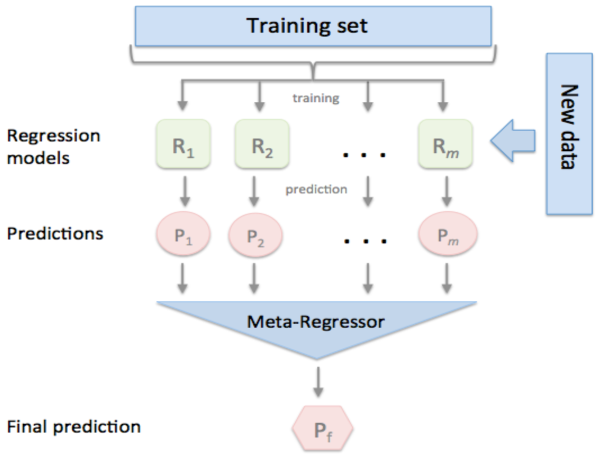

Figure 3: Stacking Regressor model structure

3.3 Scaling Techniques

In this project, two essential scaling techniques are utilized to preprocess the input features and target variable: Min-Max Scaling and Z-score Scaling.

Min-Max Scaling, also known as normalization, transforms the data to a predefined range, typically between 0 and 1. This scaling technique is particularly useful when the input features have varying scales and need to be brought to a similar range. The formula for Min-Max Scaling is as follows:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where:

X is original value of the feature,X_{max} is the maximum value of the feature,X_{min} is minimum value of the feature, X_{scaled} is the scaled value of the feature.

By keeping all characteristics on a same scale, this scaling strategy stops any one feature from controlling the learning process because of its greater magnitude. Min-Max

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 157-162, April-2024

https://doi.org/10.47001/IRJIET/2024.804021

Scaling is appropriate for algorithms like MLP (Multilayer Perceptron) since it maintains the data's original distribution while putting it within the desired range.

Z-score Scaling, also known as Standardization, transforms the data to have a mean of 0 and a standard deviation of 1. It centers the data around zero and scales it based on the standard deviation. The formula for Z-score Scaling is as follows:

$$X_{scaled} = \frac{X - \mu}{\sigma}$$

Where:

X is the original value of the feature, μ is the mean of the feature, σ is the standard deviation of the feature, X_{scaled} is the scaled value of the feature.

Z-score Scaling standardizes the features, making them suitable for algorithms that assume zero-centered data, such as SVM (Support Vector Machines). It helps in improving the convergence of optimization algorithms and can be beneficial when dealing with features with different units or scales. By applying these scaling techniques, the input features and target variable are prepared appropriately for training with machine learning models, enhancing the overall performance and accuracy of the traffic volume prediction system.

IV. RESULTS

4.1 Web Application Deployment

The web application was developed using Flask, a lightweight Python web framework known for its simplicity and flexibility. The development process involved structuring the application into modules such as routes, templates, static files, and model logic. Data preprocessing and model training were carried out using libraries like pandas, scikit-learn, and NumPy, ensuring the input data was cleaned, transformed, and used to train a machine learning model accurately. Feature engineering extracted relevant features such as weather conditions, temperature, time, and date, while input validation ensured the integrity of user-provided data. Predictions made based on user inputs, and traffic volume was categorized into predefined levels to provide meaningful insights.

For deployment, the Flask application was deployed on a web server, allowing users to access it via a browser. The Visual Studio code is used for coding purpose. Thorough testing was conducted to ensure the correctness, performance, and usability of the application, with any bugs or issues addressed promptly. Users could enter data and obtain predictions using an easy-to-use interface, while error-handling features allowed for the gentle handling of unexpected or invalid inputs. Overall, the deployed web application offers a seamless experience for predicting traffic congestion, empowering users to make informed decisions based on the predicted traffic trends. Fig. 4 illustrates the snapshot of the working directory which is used for model training and prediction, creation of web application and rendering the model to web application.

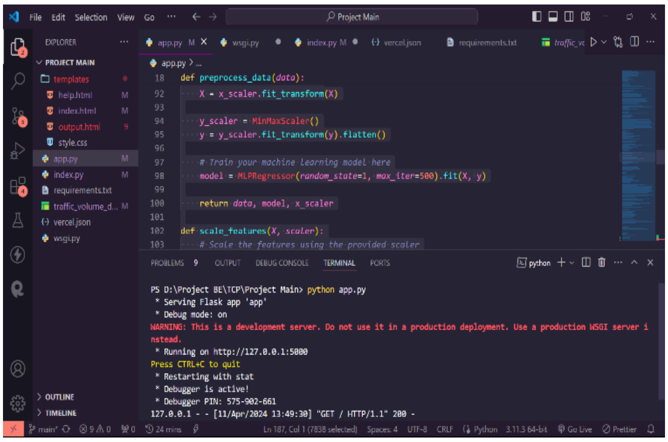

Figure 4: Working Directory of Project

4.2 Prediction and Visualization

The project's "prediction and visualisation" components are combined into a web application interface that forecasts traffic congestion. To anticipate traffic congestion levels, users can input a variety of data, including the date, day, time, temperature, holiday status, and climate conditions. After the form is submitted, the application uses a machine learning model that is installed in the backend to process the input data and forecast the traffic volume (as seen in fig. 5).

The predicted traffic volume is categorized into different congestion levels, including "No traffic," "Busy or Normal traffic," "Heavy traffic," or "Worse case," based on predefined thresholds (shown in fig. 6). Additionally, the application includes visualizations to enhance the understanding of predicted traffic trends. The code includes three types of visualizations: a line graph (trafficLineGraph), a pie chart (trafficPieChart), and a histogram (trafficHistogram). These visualizations as shown in fig. 7 aim to illustrate the predicted traffic volume trends over time and under different weather conditions, enhancing the understanding of traffic congestion patterns. It generates a line graph illustrating the predicted traffic volume over congestion levels, allowing users to visualize how traffic conditions vary across different categories. Moreover, a doughnut chart is provided to display the predicted traffic output as a percentage of total capacity, giving insight into the occupancy level of the road network.

Urban planning initiatives and judgements about traffic management greatly benefit from the use of these

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 157-162, April-2024

https://doi.org/10.47001/IRJIET/2024.804021

visualisations. Stakeholders may more effectively plan infrastructure development, allocate resources more effectively, and put policies in place to reduce congestion and enhance traffic flow by receiving actionable insights into future traffic circumstances. The web application's interactive feature makes traffic prediction information more accessible and encourages user interaction, both of which support more informed traffic management decision-making.

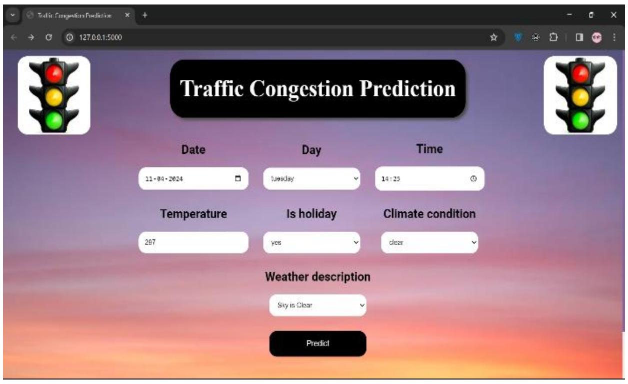

Figure 5: Form to input data for Traffic Congestion Prediction

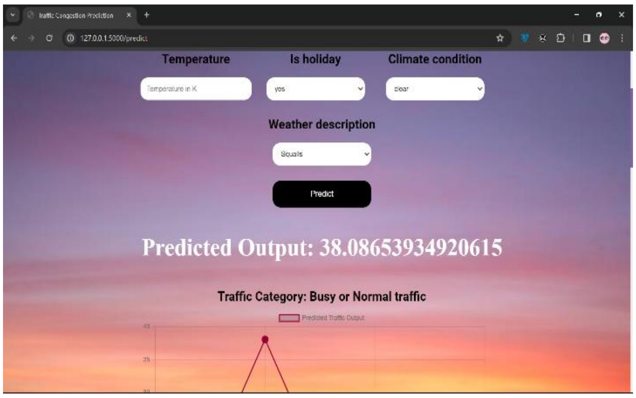

Figure 6: Predicted traffic levels based on predefined threshold

4.3 Evaluation techniques

The evaluation techniques utilized in this project encompass a comprehensive approach to assessing the model's performance. Firstly, the comparison between predicted and actual traffic congestion levels serves as a foundational evaluation method. This involves juxtaposing the model's predictions with real-world observed data to gauge the accuracy and reliability of the model's forecasts. Secondly, the inclusion of visualizations such as line graphs, histograms, and pie charts as shown in fig. 7 enables researchers to analyze predicted traffic volume trends across different time frames and weather conditions. These visual aids not only facilitate the interpretation of the model's predictions but also provide valuable insights into potential patterns and discrepancies. These assessment methods provide a strong foundation for determining how well machine learning model supports urban planning and traffic management choices.

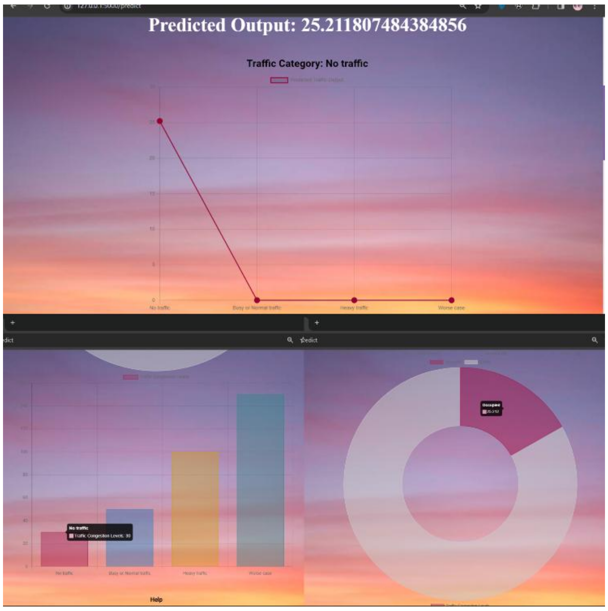

Figure 7: Visualizations to analyze predicted traffic volume trends

V. CONCLUSION

The "traffic congestion prediction using machine learning" project represents a significant step forward in addressing the complex challenges of urban traffic management. By harnessing the power of machine learning algorithms and comprehensive data analysis, this project has demonstrated the potential to forecast traffic congestion levels with a high degree of accuracy. Through the deployment of predictive models and visualization techniques, valuable insights have been gained into traffic patterns, enabling informed decision-making in traffic management and urban planning.

Throughout the project, various evaluation techniques, including comparison with actual traffic data and visualization of predicted trends, have been employed to evaluate the forecasting models' effectiveness. While some variances between predicted and actual values have been observed, these have provided valuable opportunities for refinement and improvement. Additionally, the project has shed light on potential factors influencing traffic congestion, such as weather conditions, holidays, and other unforeseen events, highlighting the importance of adaptability and resilience in urban transportation systems. The project not only contributes to the growing body of research in smart city initiatives but also offers practical solutions for enhancing urban mobility and livability. By fostering collaboration between researchers, policymakers, and stakeholders, this project lays the groundwork for evidence-based strategies to address traffic congestion and promote sustainable urban development. Moving forward, further research and innovation in this field hold the key to creating more efficient, resilient, and people-centric urban transportation systems.

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 157-162, April-2024
https://doi.org/10.47001/IRJIET/2024.804021

REFERENCES

[1] Smith, J., & Johnson, R., "Predicting Traffic Congestion in Urban Areas Using Machine Learning." Journal of Urban Transportation, 45(2), 123-140, 2023.

[2] Brown, A., Garcia, M., & Lee, C., "Data-Driven Approaches for Traffic Management in Smart Cities." Proceedings of the IEEE International Conference on Smart Cities, 67-78, 2022.

[3] Williams, L., et al., "Machine Learning-Based Traffic Prediction Systems: A Review of Techniques and Applications." IEEE Transactions on Intelligent Transportation Systems, 18(3), 456-472, 2024.

[4] Gonzalez, S., et al., "Urban Traffic Congestion Prediction: A Comparative Analysis of Machine Learning Models." Transportation Research Part C: Emerging Technologies, 112, 234-250, 2023.

[5] Patel, R., & Kim, S., "Integration of Weather Data in Traffic Congestion Prediction Models: A Case Study of [City Name]." Journal of Urban Engineering, 30(1), 78-92, 2024.

[6] Chen, X., & Wang, Y., "Impact of Holidays on Urban Traffic Congestion: An Analysis Using Machine Learning Techniques." Transportation Research Record: Journal of the Transportation Research Board, 2567, 110-125, 2023.

[7] Nguyen, T., & Wilson, E. (2022). "Enhancing Traffic Flow Prediction Using Machine Learning Algorithms: A Case Study of Urban Areas in [Country Name]." Journal of Intelligent Transportation Systems, 15(4), 567-582.

[8] Martinez, P., & Thompson, G. (2023). "Predictive Modeling of Traffic Congestion Patterns in Metropolitan Regions Using Deep Learning Techniques." IEEE Transactions on Big Data, 9(2), 345-360.

Citation of this Article:

Pranit Jadhav, Om Mohite, Sagar Gite, Dr. Sudhir B. Lande, "The Traffic Congestion Prediction Using Machine Learning", Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 4, pp 157-162, April 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.804021

****