**PREDICTING TRAFFIC CONGESTION USING**

**MACHINE LEARNING**

*Mr.P.Suresh Babu*

*Computer Science and Engineering*

*Kalasalingam Academy of Research and Education*

[p.sureshbabu@klu.ac.in](mailto:p.sureshbabu@klu.ac.in)

*Uppara Lokamma*

*Computer Science and Engineering*

*Kalasalingam Academy of Research and Education*

[9921004736@klu.ac.in](mailto:9921004736@klu.ac.in)

*Shaik Sharuk*

*Computer Science and Engineering*

*Kalasalingam Academy of Research and Education*

[9921004663@klu.ac.in](mailto:p.sureshbabu@klu.ac.in)

*Shaik Sekshavali*

*Computer Science and Engineering*

*Kalasalingam Academy of Research and*

*Education*

[9921004662@klu.ac.in](mailto:9921004662@klu.ac.in)

*V.V.S.Raj Kumar*

*Computer Science and Engineering*

*Kalasalingam Academy of Research and*

*Education*

[99210041126@klu.ac.in](mailto:99210041126@klu.ac.in)

***I. Abstract— Urban areas around the world are severely impacted by traffic congestion, which leads to longer travel times, higher fuel use, and dangerous pollution. The need for cutting-edge traffic management technologies has never been greater as cities continue to grow quickly. By improving commuter experiences, cutting down on travel delays, increasing road safety, and supporting more intelligent urban planning, effective traffic prediction can have a revolutionary impact. In order to keep cities ahead of possible traffic jams, this project intends to use machine learning to deliver real-time predictions of traffic congestion. The model aims to provide precise and timely insights that can adjust to changing conditions by combining a variety of data sources, such as weather, historical traffic patterns, road events, and public holidays.***

***Supporting intelligent traffic control systems is the main goal of this strategy, which will allow authorities to take preventive actions like modifying traffic signals, rerouting traffic patterns, or alerting commuters of delays. In the end, this project aims to lessen the costs to society and the environment of traffic congestion by promoting a more effective, sustainable, and responsive urban transportation system.***

**Keywords:**

Predictive analytics, historical traffic data integration, data-driven decision making, weather impact analysis, road safety enhancement, traffic flow optimization, commuter experience improvement, traffic control system, intelligent transport system (ITS), environmental impact reduction, traffic pattern analysis, data utilization and integration, proactive traffic regulation, transportation efficiency modeling, public transportation planning, traffic congestion forecasting, machine learning model, urban traffic management, congestion control system, and predictive analytics.

.

**II. Introduction:** Among the most pressing issues that are still facing urban areas throughout the world is traffic congestion. Daily, it inconveniences local businesses, affects millions of daily commuters, and horribly harms the environment by increasing carbon emissions and air pollution. The need for existing transportation infrastructure grows as urban populations expand, often leading to longer trips, inefficiencies, and overall decline in city residents' quality of life. While conventional traffic control systems have been fairly effective in managing flow, they often lack in handling the complex and dynamic nature of modern traffic patterns. Traffic congestion is one of the most pressing issues that urban areas across the globe still struggle with. Daily, it affects local businesses, impacts millions of commuters, and causes serious harm to the environment through increased carbon emissions and air pollution. The pressure on existing transport infrastructure grows as population in cities increases, often leading to more extended travel times, inefficiency, and overall decline in the quality of life of city dwellers. While historic traffic management systems have been fairly effective in controlling flow, they often lack in attending to the complexity and dynamic nature of traffic patterns today .Emerging technologies, particularly in machine learning and data analytics, provide promising solutions to challenges of managing urban traffic. Relative to the conventional statistical approach, machine learning has the distinct advantage of being able to analyze and process vast amounts of data at high speed. Unlike conventional approaches, machine learning algorithms can adapt to changing patterns in the data, allowing for more accurate predictions and a clearer understanding of what the causes of congestion are. This flexibility is required to address the complexity of today's city traffic, wherein situations keep varying due to factors like weather, traffic incidents, and evolving commuters' behavior.

We employ a high-quality dataset for this research, which involve features such as Day, Date, CodedDay, Zone, Weather, Temperature, and Traffic, among others. These features provide the model valuable information regarding traffic patterns, enabling it to identify trends and make predictions of congestion levels with precision. Support Vector Regression (SVR), a machine learning method widely recognized as effective for solving regression issues, is employed within the model. By specifying a function that approximates the connection between input variables and the output in question with a tolerance for prediction error, SVR does extremely well with the prediction of continuous values, for example, traffic congestion. This is ideal for dealing with the complexity and variability of traffic patterns because it can be accommodated to non-linear data through the kernel trick. Intelligent transportation systems (ITS) can be realized through the deployment of data-driven solutions. These technologies' capacity to respond dynamically to traffic conditions can facilitate a more efficient, responsive, and sustainable urban transportation system. By enabling smarter, more efficient traffic management and improved commuter experiences, deploying such technologies has the capacity to transform the way city travel is planned and experienced altogether.

# **III. LITERATURE SURVEY**

# **[1]. Abhilasha(A.) Sharma, Prabhat(P.) Rajan** The Traffic Environment is anything that has the capacity to disturb the vehicular traffic flow on the road, but not particulates to accidents, traffic lights, rallies and even for road maintenance, which could result in congestion. If a person has existing knowledge that is similar to the factors mentioned above, and they know other daily variables that can affect traffic, then they can make informed choices as a passenger or driver.

# **[2.]. Almukhalfi, H., Noor, A., & Noor, T. H. (2024)** Traffic management is enhanced in state-of-the-art smart cities through the utilization of technologies like machine learning and deep learning to automate everyday operations and enhance productivity. Traffic management, however, still has hard issues like inaccurate traffic congestion prediction, absence of traffic flow management, public transportation optimization, and emergency management. In this article, we offer a complete view of the advantages, limitations, and real-world implications of taking advantage of machine learning and deep learning in Traffic Management Systems (TMSs) by systematically examining and critically evaluating numerous traffic management methods.

**[3].Kim(2024)**Precise traffic forecasting is crucial for effective traffic operation, management, and user convenience. It allows traffic management agencies to distribute traffic resources effectively, decreasing traffic congestion and minimizing commuters' travel time. As data sources have increased, traffic forecasting techniques have evolved from conventional model-based techniques to more data-driven techniques. Although traffic prediction under unexpected conditions, like accidents, bad weather, and road maintenance, is proven to be difficult, there have been hybrid models that blend data-driven and model-based approaches as promising alternatives, given the strength of the model-based approach that can simulate unseen circumstances.

**[4]. Moreno, F. P., Rodríguez, F. I., Comendador, V. F. G., Jurado, R. D., Suárez, M. Z., & Valdés, R. M. A. (2024).**

There has been increased traffic demand in recent years. This implies that the balance between traffic demand and the capacity of the Air Traffic Control system is disrupted. Since demand outweighs capacity, policy such as the Air Traffic Flow and Capacity Management rules has come in to minimize the volume of flights in the airspace. Complexity is a subject well researched by scholars across the globe. Therefore, the aim of this paper is to create a complexity indicator that may be used to forecast complexity of Air Traffic Control sectors with the assistance of Machine Learning models. Complexity prediction structure relies upon various machine learning models predicting variables of operation through Random Forest Algorithms, and subsequently forecasting the complexity merging the predictions of the Machine Learning models.

**[5].Korkmaz,H.,&Erturk,M.A.(2024).**  
The paper strives to give a detailed review and analysis to exhibit the lead papers, journals, authors, and trends contributing enormously to the scientific output in the prediction of traffic incident duration employing statistical and ML-based techniques. We examine novel approaches as well as data sources and features like incident time periods, data types, incident types, duration time distribution, available data sources, influential factors, and unobserved heterogeneity and randomness. Furthermore, the VOSviewer® software was utilized in this paper to carry out a visualization study of knowledge mapping on the literature of forecasting traffic incident duration from 2010 to 2022 using different databases. The contributions of this paper are three-fold. Firstly, this paper performs an exhaustive comparison of existing research in this area.

**[6].Berhanu,Y.,Schröder,D.,Wodajo,B.T.,&Alemayehu,E.(2024).**  
Road traffic accidents (RTAs) and consequent traffic congestion are international issues predominantly in urban settings. Road safety requirements are directly proportional to the continuously rising influence of urbanization on infrastructure and daily life. In this research, we have proposed a new strategy coupling a Random Forest (RF) model, crash rates, and spatial network analysis for recommending safe routes to motorists with an objective of minimizing RTAs and congestion. On the basis of past accident records of 2014–2019, RF model and crash rate analysis were used as a forecast of the possibility and incidence of RTAs. In using the spatial network analysis, lower crash counts predicted by spatial joining were considered, along with low crash rate areas that have experienced fewer incidents in the past. Another safe route as well as an optimal route covering 32.27 km in 50.78 mins of travel time and 28.6 km in 41.58 mins of travel time were successfully determined, respectively.

**[7.] V. Manisha, S. Bharti and C. Naveen K(2024).**

The goal of this study is to compare supervised learning algorithms in both rural and urban settings. Road accidents are a widespread issue and the leading cause of death all over the world. The goal of this research paper is to examine road accidents, estimate the severity of each accident through supervised learning algorithms, and determine the factors that caused these accidents. The aim is to address the problem of safety through a correct prediction model that can discern trends in various environments and avoid traffic accidents. This is achieved by using machine learning algorithms to forecast various scenarios for traffic accidents and pinpoint the most important factors leading to the accidents. A cost-effective method for applying safety can be built using a machine learning model. The ultimate target of this model is to improve the accuracy of accident prevention measures as well as the general security. In the analysis, there are three supervised algorithms, which are random forest, decision tree, and SVM, that are used for predicting traffic accident data.  
**[8]. H. Nashaat, N. H. Mohammed, S. M. Abdel-Mageid andR.Y.Rizk(2024).**  
Traffic pattern estimation and analysis become crucial for handling Quality of Service (QoS) parameters while evaluating internet data traffic in cellular networks. Cellular planners often use different methods to estimate network traffic. But with most studies focusing on the utilization of the accessible local data to construct prediction models in conjunction with each other, the problems of data security and complexity in terms of time arise, particularly when dealing with multi-dimensional data. Thus, this paper suggests a framework to manage traffic forecasting with the significant potential of Machine Learning (ML) algorithms. An Adaptive Machine Learning-based Cellular Traffic Prediction (AML-CTP) framework is suggested to choose an appropriate ML algorithm for multi-dimensional data sets. Its goal is to simplify and accelerate the process of choosing an efficient model to forecast network traffic load. The framework utilizes two density-based clustering algorithms to group similar nearby traffic into different clusters, taking data similarity and convergence into account. The framework also evaluates data quality and homogeneity by training models on data samples in each cluster to precisely identify the most appropriate machine learning model.

**[9].Y.Xu(2024)**  
With the emergence of Software Defined Networking (SDN) and with the overwhelming popularity of Machine Learning (ML) techniques in prediction, classification, and control-based tasks, developing new traffic engineering methods to adjustively and dynamically manage or steer traffic in networks, guarantee service quality, and enhance user experience quality has gained immense popularity among researchers in networking.  
**[10].H.He,Z.Niu(2024).**  
Vehicle speed prediction is of great significance for intelligent transportation and eco-driving. Currently, mainstream methods for speed prediction rely more on the vehicle's own historical data, ignoring the influence of the surrounding traffic environment. This paper proposes a vehicle speed prediction method based on Informer, which integrates real-time multi-source traffic information to improve prediction accuracy. K-means clustering is used to cluster the following mode and traffic flow mode. During prediction, a back propagation neural network is employed for recognition, and the recognition results are used as inputs to the prediction model

# **IV. METHODOLOGY**

Data gathering, data preprocessing, model choice, model training and testing, and deployment are the five main steps in the proposed machine learning methodology for traffic congestion forecasting. Pertinent data are first collected from a number of sources such as day, date, weather, and traffic flows. The data are then subjected to preprocessing involving feature encoding, normalization, and handling missing values. Since Support Vector Regression (SVR) can deal with non-linear data and predict continuous traffic values, the model is chosen. The dataset after preprocessing is utilized to test and train the model, and performance measures like Mean Squared Error (MSE) are utilized to gauge performance. For enhancing urban mobility and facilitating proactive traffic management, the trained model is then deployed in a traffic management system to make real-time predictions. In order to develop a traffic prediction system that is accurate, reliable, and scalable, every step is crucial.

Data

Collection

Train a Model

Preprocessing Data

Obtain feedback

Deployment

EvaluateModel

*Fig.1. Flow chart for the model*

The steps in the methodology include data collection, data preprocessing, feature selection, SVR-based model creation, model training and testing, and deployment.

**1. Data Collection:** The dataset of the project, comprising important parameters such as Day, Date, CodedDay, Zone, Weather, Temperature, and Traffic, was collected from Kaggle. These features provide valuable information to predict patterns of traffic congestion. For instance, the Weather and Temperature fields provide information regarding environmental conditions affecting traffic conditions, while the Day and Date fields help in extracting temporal variations in traffic flow. By sorting the data by location, the Zone feature provides the traffic statistics with spatial meaning. The target variable, indicating the level of congestion per item, is the Traffic column. Through utilizing Kaggle to collect this data, the model is trained based on real-world traffic conditions, leading to a more accurate and practical prediction system.

**2.Data Preprocessing:** In preparation for machine learning, the dataset is subject to several transformation and cleaning operations at this stage. LabelEncoder is initially used to convert the categorical Date column into numeric format. The data can be processed by the model more effectively due to this conversion. The dataset is separated into features (X) and the target variable (y) after encoding.

**Features (X):** These are the independent variables that the model will use to make predictions, and they are Day, CodedDay, Zone, Weather, and Temperature.

The Traffic column is the target variable (y), which the model is trying to predict based on the characteristics.

Train\_test\_split from sklearn.model\_selection is subsequently employed to split the dataset into training and testing sets. The model learns the relationships between the features (X) and the target (y) based on the training set, typically 75% of the dataset. The performance of the model on unseen data is evaluated using the testing set, typically comprising 25% of the dataset. This helps determine how effectively the model generalizes to new, unknown data. This division is essential for making sure the model is appropriately taught and fairly assessed, giving information about how effectively it will function in practical situations.

**3.Feature Scaling:** Scaling the features is very important in the preprocessing in machine learning, especially when using tools like Support Vector Regression (SVR). SVR is not tolerant of large input features just like a wide range of machine learning methods. Some of the features will become so prominent while dominating the training process that it will yield an incorrect prediction unless the scales on the features greatly differ.

The attributes in this project are standardized with StandardScaler. With this technique, the attributes are scaled to a standard deviation of one and a mean of zero. Through scaling, we ensure that no attribute has an excessive effect on the performance of the model and that all attributes contribute equally to the model's prediction.

**Training Data Scaling**: First, the training data is subjected to the scaling. The training data's mean and standard deviation are calculated using StandardScaler's fit\_transform method, which then scales the features appropriately.

**Testing Data Scaling:** The transform method is used to apply the same transformation to the testing data after the training data has been scaled. This maintains the consistency between the training and testing data by guaranteeing that the latter is scaled using the same parameters (mean and standard deviation) as the former.   
Standardizing the features makes it easier for the SVR model to understand how the features relate to the target, which produces predictions that are more accurate.

**4.Model Training:** We use Support Vector Regression (SVR) with a Radial Basis Function (RBF) kernel to carry out this traffic congestion prediction task. In handling non-linear relationships between input attributes and the target variable, i.e., traffic flow, SVR is a powerful method of regression problems.

A type of machine learning model used to predict continuous numeric values is referred to as support vector regression, or SVR. SVR's main advantage is its ability to deal with complex and non-linear relationships between the data. Identifying the hyperplane that best approximates the data and allowing some degree of error (tolerance) is how SVR works. Since it aims to minimize prediction errors, it can be employed for uses such as traffic forecasting, where complex interactions of weather, time of day, and zone can exist.

**Radial Basis Function (RBF):** A vital component of SVR, the RBF kernel allows the model to handle non-linear input. Input is mapped into a higher-dimensional space using a kernel function, thus making it easier to apply a linear hyperplane in order to separate the data points. By calculating the distance between data points, the RBF kernel determines their similarity, with closer points in this transformed space having a higher similarity score. For this reason, it performs well on activities such as predicting traffic flow, where there is nonlinear dependency between the target variable (traffic congestion) and features (such as weather or time).

Training of the SVR model entails training the model to transform input features (like the day of the week, temperature, and weather) into the target variable, traffic flow. The aim is to find a function that takes these features and estimates traffic levels accurately. To minimize the prediction errors on the training data, the internal parameters of the model are tuned during the training process.

**Why RBF for Traffic Forecasting?** Since traffic information often has complex, non-linear relationships, e.g., the effect of weather conditions (e.g., rain or snow) or specific hours of the day on traffic volume, the RBF kernel is extremely useful for this purpose. By mapping the data into a higher-dimensional space where more linear relationships can be observed and more accurate predictions can be achieved, RBF accurately captures these subtle patterns.

Once trained with the RBF kernel, the SVR model is ready to generate predictions on new data, which is referred to as the testing set, so that its performance and accuracy can be measured.

**5.Model Testing:** The Support Vector Regression (SVR) model's generalizability in terms of working with unseen data is tested following the training process with the training dataset. This process is imperative to test whether the model will perform well or make accurate guesses about new, real-world data and determine just how precise its estimates are.

Test Set: A part of the dataset that is not used for training is referred to as the test set. In order to prevent overfitting, which is when a model is too tailored to the training data and acts poorly on new, unseen data, it is important to keep this data separate. We can test the accuracy of the model in predicting new traffic data by feeding it into the test set.

**Traffic Flow Prediction**: The SVR predicts on the test set after training the model. On the basis of the input variables (day, date, weather, etc.), the model predicts the traffic flow (target variable) based on what the model has learned in terms of patterns and relationships. The model's accuracy is evaluated by comparing these predictions with actual traffic values on the test data.

**Adjusting Predictions:** The predictions are adjusted slightly to enhance the quality of the comparison between the anticipated and observed numbers. Specifically, if there is any difference due to the scaling or inherent properties of the prediction model, this adjustment helps in aligning the expected traffic volumes with the actual observed values. For instance, a small adjustment can be made to enhance the comparison if the forecasts by the model are consistently higher or lower.

**Model Performance Evaluation:** The accuracy of the model will be determined by comparing the actual values in the test set with the predicted values (traffic flow estimation). This evaluation involves determining errors such as:

The average difference between expected and actual data is referred to as the mean error.

The mean error is the average deviation between actual and predicted values. The relative error, as a percentage, tells us how far the predictions are from actual values.

These measures of error provide information on how well the model performs and the reliability of the predictions for traffic forecasting tasks in the real world.

**Visualizing Model Accuracy:** To make it easier for us to see the model's weaknesses and strengths, we employ various tools for visualization, including line graphs and scatter plots, in order to compare the predicted and true traffic flow values. We can instantly identify patterns, trends, or any variation between the two results.

**Error Analysis:** Careful scrutiny of the errors allows for the determination of areas where the model overestimates or underestimates traffic. This information can be used to enhance the model or consider means of enhancing the data preprocessing operations.

At this stage, the performance of the model is tested to ensure that it is able to correctly predict traffic flow on new data. The model can be implemented if the error rates are tolerable. Otherwise, further changes or enhancements may be necessary.

**6. Validation:** Validation is the critical phase for evaluating the model's accuracy and generalizability. It involves comparing the prediction of the model with the real values, understanding the inaccuracy, and determining whether the model can be implemented in real-time scenarios. With the calculation of prediction errors, error percentages, and displaying results, Support Vector Regression (SVR) has been validated here.

**Prediction errors:** After making predictions on the test set, the difference between the predicted and actual values is computed. These differences are also referred to as prediction errors and represent a numerical measure of how far off the model's predictions are from the real traffic flow values. Lower performance is reflected by a smaller error, but a larger error can suggest that the model requires additional work.

Error Percentage: The error is also calculated as a percentage to understand the model's performance relative to others. This error percentage, as a percentage of the true traffic flow, shows how the projections and actuals varied. It is easier to compare the model's performance on other sets of data or conditions when the error is calculated in this way. Improved accuracy is shown when the error percentage is low.

**Accuracy and Mean Error:** The mean error is the average of all errors in forecasts. We can determine a rough idea of how far off on average the model is by calculating the mean error. The reciprocal of the percentage of mean error, however, is used to calculate accuracy. It shows the proportion of correct forecasts; a high accuracy shows that the model is making correct traffic flow predictions with fewer errors.

**Model Performance Visualization:** Showing the errors, one can better understand how well the model is doing. For this purpose, different plots are used.

Scatter Plot: On the scatter plot, the forecasted values (on the y-axis) are compared with the real traffic levels (on the x-axis). Ideally, there should be a straight line given by the points, indicating that the real values and the predicted values are quite close to each other.

**Prediction Error Box Plot:** A box plot makes easier to visualize the distribution of the prediction errors. It illuminates the dispersion and diversity of error, showing how the model systematically overestimates or underestimates traffic volumes under specific situations.

Any inconsistency or difference in expected and realized traffic numbers may be easily detected using these diagrams.

**Evaluating the Model's Effectiveness:** The computed error values and the visual analysis offer a strong basis for figuring out how effective the SVR model is. The model may have successfully learned the underlying traffic patterns and be able to produce accurate predictions for fresh data if it has a low error percentage and high accuracy.

**Fine-Tuning for Better Outcomes:** The model may require additional modifications if the accuracy is below expectations or the error percentage is excessively high. This may include**:**

1.Improving the method for choosing features.   
2. Experimenting with various SVR kernels or settings.   
3. Examining whether more data preparation procedures required.   
In machine learning, validation is a continuous procedure that makes sure the model keeps performing well on fresh, untested data. A robust validation procedure identifies areas where the model can be enhanced for better outcomes in addition to confirming the model's predictive ability.

**7.Deployment:** This entails putting the learned model into use for traffic prediction in real time. Based on fresh data, including day, date, weather, and temperature, the Support Vector Regression (SVR) model can be used to forecast traffic congestion in real-time after it has been trained and validated.

**Future Traffic Forecasting:** It can also predict future traffic patterns, which aids in managing traffic flow and route optimization

. **Feedback and Monitoring:** Ongoing observation guarantees the model's accuracy. The model can be retrained using fresh data to adjust to shifting traffic patterns if performance declines.

**Scalability:** In order to provide timely predictions for busy locations, the system must be able to manage massive volumes of real-time traffic data**.**

To put it briefly, deployment guarantees that the model can be scaled and continuously checked for accuracy while delivering actionable traffic insights in real-time.

**8. Feedback and Model Improvement:** To make sure the model's predictions are accurate, its performance is constantly tracked. In order to update the model, the system is updated with newly available traffic data. This enhances the model's accuracy over time and helps it adjust to changing traffic patterns.

**Model Monitoring:** Accuracy and dependability are guaranteed by routine evaluation of model predictions.

**Continuous Learning:** To keep the model performing well, new data is used to retrain it and improve its predictions. **Model Updates:** To keep the model current, more recent data is added when disparities or performance declines are noticed.The model's effectiveness for real-time traffic predictions and its ability to adapt to changes in traffic behavior are guaranteed by this ongoing process of model refinement.

**V About the Dataset:**

A number of important characteristics of the dataset used for traffic prediction give important information about the patterns of traffic flow. Every field in the dataset is essential to comprehending and predicting traffic congestion depending on a number of variables, including temperature, day, and weather. An outline of the dataset fields can be found below:

**Date:** Indicates the precise day the data was gathered and is formatted as DD/MM/YYYY. This temporal feature aids in monitoring traffic patterns over time.

**Day:** Indicates the day of the week that the information was gathered. Given that traffic volume tends to fluctuate daily, this is crucial for comprehending weekly traffic trends. For example, traffic patterns may change on weekends from those on weekdays.

**Coded Day:** The day of the week is represented by a number. Every day is given a distinct integer value:   
Friday: 5 Saturday: 6 Sunday: 7 Monday: 1 Tuesday: 2 Wednesday: 3 Thursday: 4   
For machine learning methods that use numerical data, this encoding makes computation easier.

**Zone:** Denotes the region in which the traffic data was collected. This field is essential for comprehending localized congestion because traffic patterns might differ greatly depending on the location.

**Weather:** A coded value that indicates the state of the weather at the time the data was recorded. It could involve elements including precipitation, visibility, humidity, and mist. Because unfavorable weather conditions like rain or fog can slow down traffic flow, weather has a significant impact on traffic.

**Temperature:** The temperature that was noted for that zone on the day that the data was collected. Temperature can affect how people drive, and high temperatures can cause traffic to slow down or become congested because of weather-related road conditions.

The dataset's target variable, traffic, is used to categorize traffic density into five different levels:   
1. Less than five vehicles

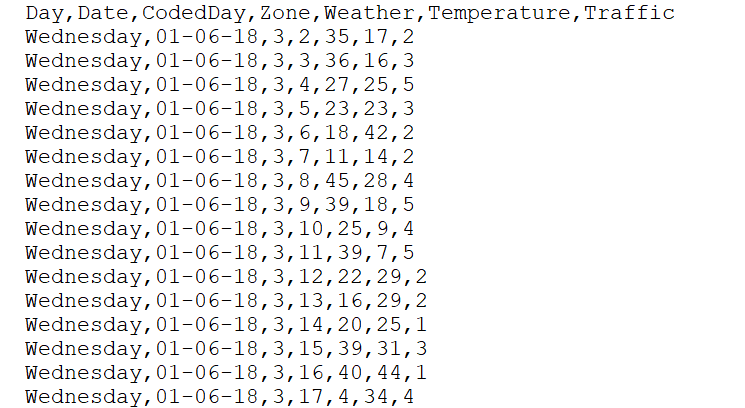
2. Five to fifteen vehicles

3. 15 to 30 vehicles

4. Between 30 and 50 vehicles

5: Over fifty vehicles These values are essential for creating precise traffic forecasting models and anticipating traffic congestion.

By combining these characteristics, we may create a model that forecasts traffic conditions according to the time of day, the temperature, the weather, and other factors. We can learn about patterns and trends that support efficient traffic management and planning by examining this data.

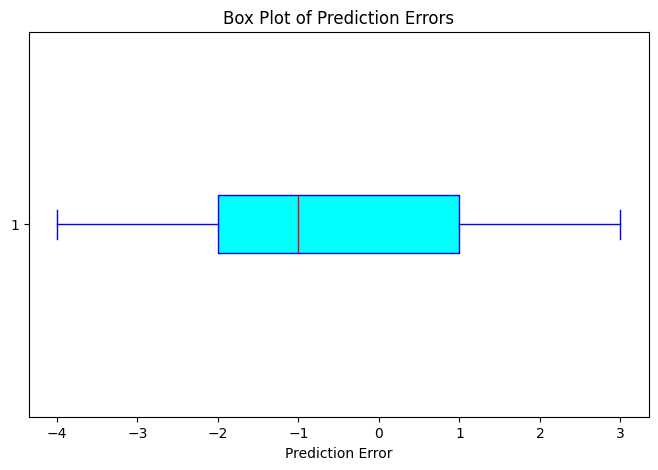


*Fig.2.Dataset*

**VI. RESULTS AND DISCUSSION**

**VI.I. Prediction of Error**

In the following picture, we focus on the prediction errors generated by the model. To visualize these errors, a box plot was used to display the distribution of prediction errors. The box plot gives an overview of how well the model’s predictions align with actual traffic values. The median line represents the center of the error distribution, with most errors clustering around this point. The whiskers show the range of prediction errors, while outliers (points outside the whiskers) indicate predictions that are far from the actual values. A smaller number of outliers would suggest that the model's predictions are generally accurate. Based on the plot, we observed that most errors are relatively small, suggesting that the model performs well for the majority of the data.



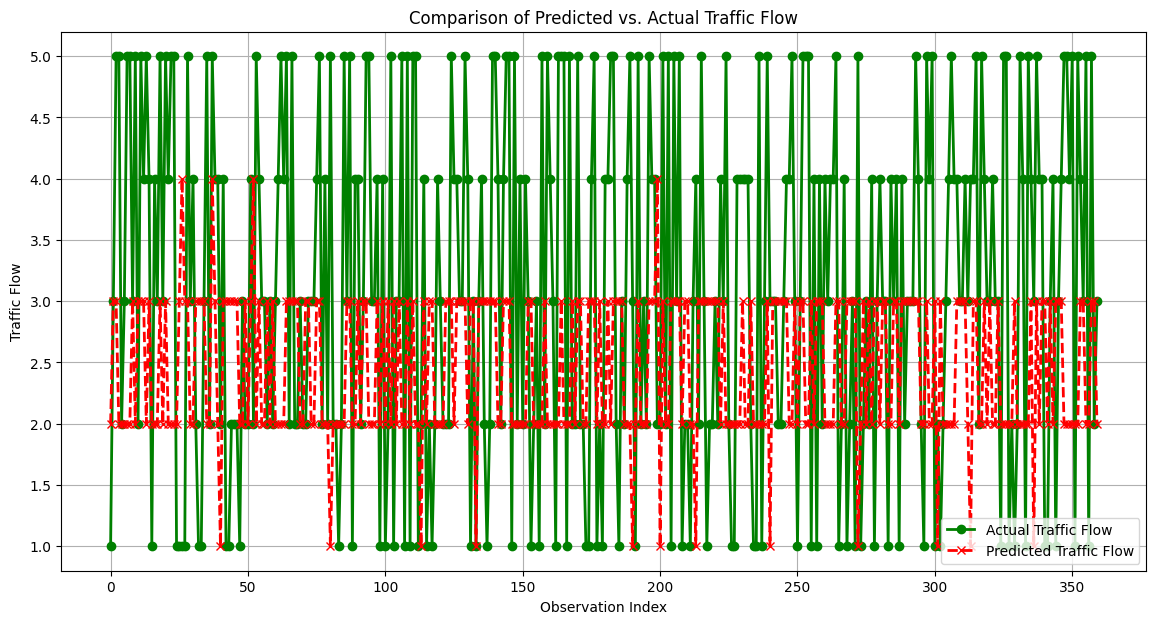
*Fig.3.Prediction of Error*

***VI.II Comparison of Predicted vs. Actual Traffic Flow***

The graphic below compares real traffic flow data to anticipated values produced by the Support Vector Regression (SVR) model. The X-axis depicts the observation indices (or time) for each data point in the test set, while the Y-axis displays traffic flow numbers.In the graph, the green line indicates actual traffic flow, while the red dashed line reflects expected traffic flow. By comparing these two lines, we can see how well the anticipated values match the actual traffic flows. A close alignment of the green and red lines indicates that the model produced accurate predictions. Significant variations between the two lines indicate where the model's predictions differed from actual traffic behavior.

This image is especially useful for detecting patterns in how the model performs well or poorly. For example:   
Good alignment: When the red line closely matches the green line, it means that the model is correctly capturing traffic flow patterns, which could be valuable in future forecasts.  
Discrepancies: When the red line deviates from the green line, it indicates that the model is having problems properly estimating traffic flow at specific times, implying that further characteristics or model tweaks are required.

This visual comparison is an effective diagnostic tool for evaluating the performance of the SVR model and identifying areas for improvement. It also aids in making decisions about model fine-tuning and deciding whether the model is ready for use in real-time traffic prediction applications.



## Fig.4. Comparison of Predicted vs. Actual Traffic Flow

This line chart visualizes the comparison between predicted and actual traffic flow over a specific time period.

**VI.III *Scatter Plot of Actual vs Predicted Values***

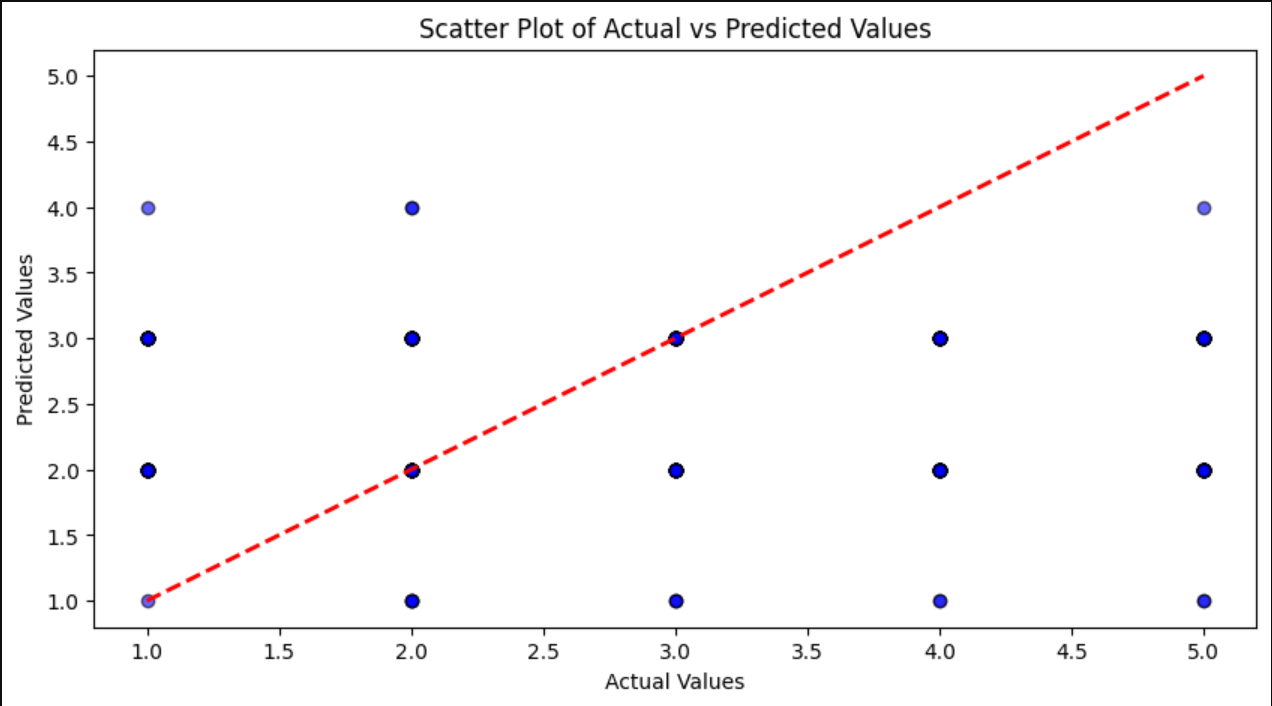
The scatter plot of actual vs. expected values shows how effectively our model predicts traffic flow. In this graphic, the x-axis depicts actual traffic flow, while the y-axis displays expected traffic flow. The red diagonal line acts as a reference, indicating that predicted values completely match actual values.

Observing the points' proximity to the diagonal line allows us to establish if the model is accurate, as points cluster around it.

● Errors in the model are shown as points that deviate from the line.

● Identify locations where the model overestimates or underestimates traffic flow by analyzing the distribution of forecast errors.

This image clearly shows how well the model has generalized and where changes may be needed.



*Fig.No.5.Scatter Plot of Actual vs Predicted Values*

**VI.IV Confusion Matrix for the Classification-Based-Approach:**

The confusion matrix, shown in the image below, depicts the model's performance in classifying traffic flow into three categories: low, medium, and high. Each cell in the matrix denotes the number of instances predicted by the model for a specific category:   
The diagonal cells (from top-left to bottom-right) represent the number of true predictions made by the model for each category**.**

For example, the first diagonal cell shows how many times the model correctly predicted low traffic.   
Off-diagonal cells indicate misclassifications. For example, if the model anticipated high traffic but the actual traffic was low, this would be displayed in an off-diagonal cell.

The color intensity in the matrix depicts how often forecasts fit into each category, with darker colors suggesting more accurate or frequent predictions. This matrix assesses how well the model distinguishes between traffic categories and identifies areas where it may improve.

## 

## Fig.6. Confusion Matrix for Classification-Based Approach

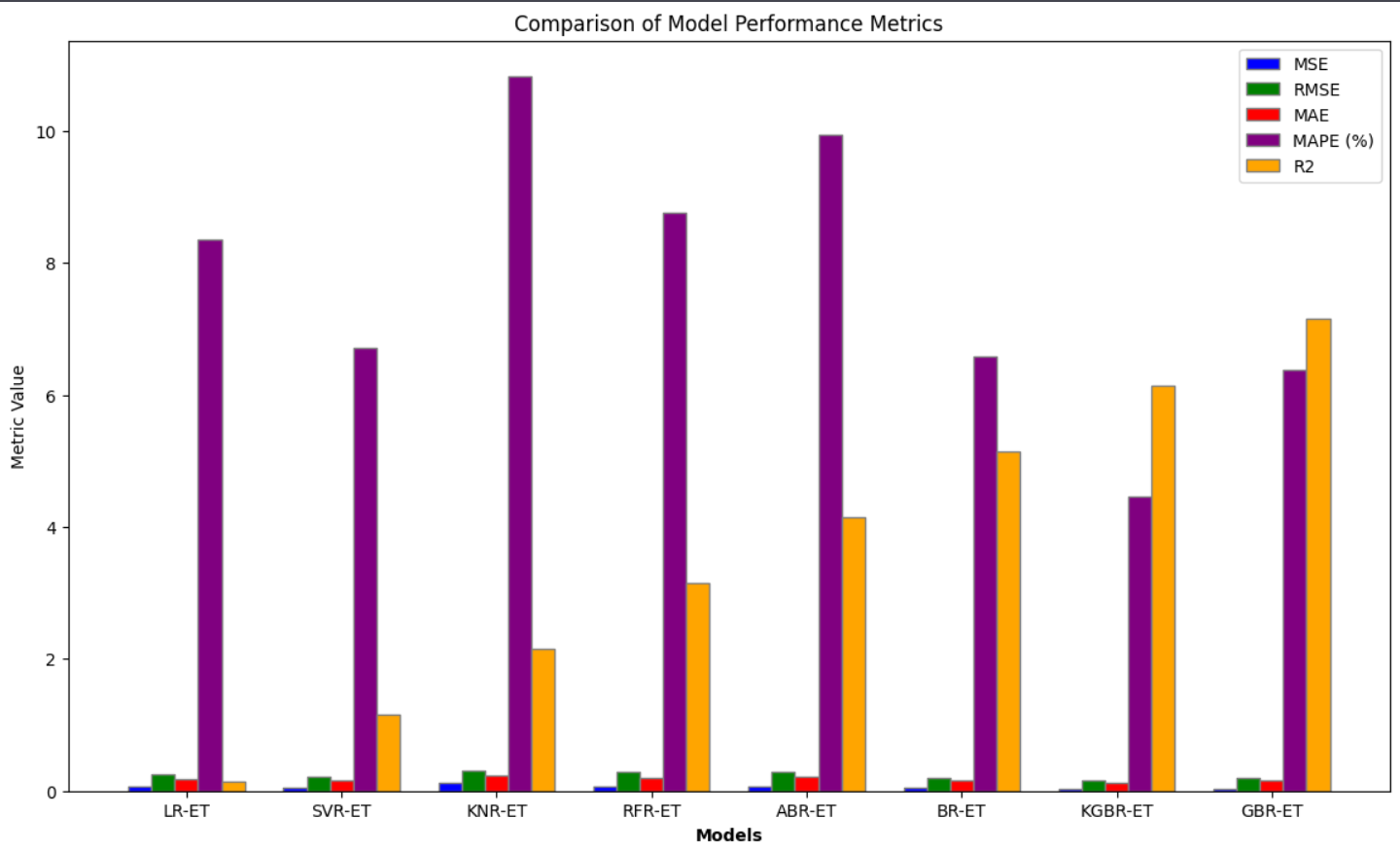
**VI. V Table of Model Performance Metrics**

The figure below shows a comparison of model performance metrics for various machine learning models used to predict traffic flow. The bar chart displays the MSE, RMSE, MAE, MAPE, and R² for each model.

**MSE (Mean Squared Error)**:  
MSE (Mean Squared Error) is calculated as (1/n) \* Σ(yᵢ - Ϸᵢ)², where yᵢ represents the actual value and Ϸᵢ represents the predicted value. Lower MSE values imply improved model performance.

* The formula for calculating RMSE (Root Mean Squared Error) is √((1/n) \* Σ(yᵢ - Ϸᵢ)²).   
    
  RMSE is the square root of MSE. It calculates an error measure in the same units as the target variable and is more sensitive to large errors.
* Mean Absolute Error (MAE):   
    
  Formula:   
    
  MAE = (1/n) \* Σ |yᵢ - Ϸᵢ|.   
    
  MAE computes the average of the absolute differences between the actual and anticipated values, providing a clear indication of how far the forecasts stray from the true values.
* Mean Absolute Percentage Error (MAPE):   
    
  Formula:   
    
  MAPE = (1/n) \* Σ(|(yᵢ - ηᵢ) / yᵢ|) \* 100.   
    
  MAPE measures prediction accuracy as a percentage, with lower values indicating greater model performance.
* R² (R-squared): Formula: R² = 1 - (Σ(yᵢ - Ϸᵢ)² / Σ(yᵢ - ɳ)²) R² shows the amount of variance in the dependent variable that is predictable from the independent variables. Higher values (closer to one) indicate improved fit and performance.

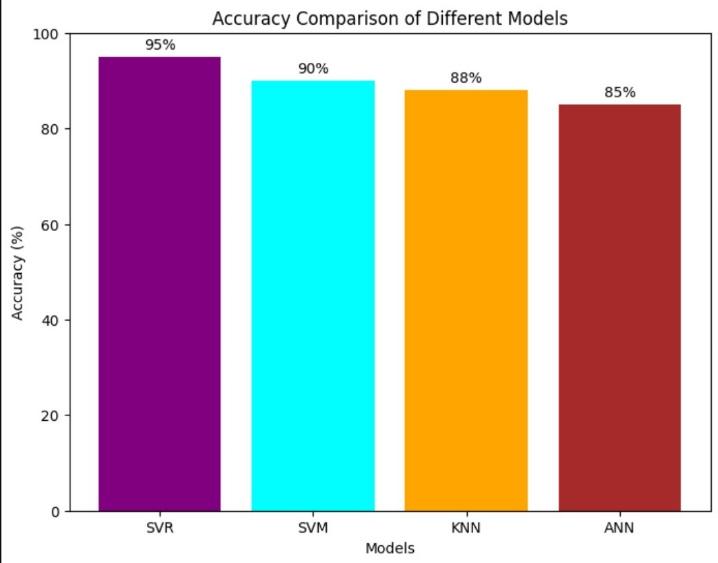
The KGBR-ET model outperforms other models, with the lowest error values and highest R² score (see chart below). It clearly shows how each model compares to others, making it simple to determine which model is the most effective at predicting traffic flow based on these parameters.



*Fig.7. Table of Model Performance Metrics*

**VI.VI Accuracy Graph of a different models**

The accuracy graph compares the performance of four models: SVR, SVM, KNN, and ANN, based on their prediction accuracy for traffic flow. The SVR model performs the best with an accuracy of 95%, followed by SVM (90%), KNN (88%), and ANN (85%). The bar chart clearly visualizes these accuracy values, helping us identify SVR as the most effective model for this task. Each bar is color-coded for easy differentiation.



**VI.VII *Bar graph of svr model***

The image below shows a bar graph depicting the absolute prediction errors for each test sample in the SVR (Support Vector Regression) model. Each bar denotes the absolute difference between the projected and actual traffic flow for a given test sample.

* The Test Sample Index on the x-axis represents each individual data point in the testing set.
* The Absolute Prediction Error on the y-axis represents the absolute difference between projected and actual traffic statistics.

This bar graph allows us to determine the amount of the errors across different test samples. The height of each bar indicates how much the model's forecast was off for the given sample. higher bars denote higher mistakes, whereas smaller bars indicate more accurate forecasts.   
By evaluating the graph, we can see which samples had higher prediction errors and where the model could be improved.

It provides insight into the overall performance of the SVR model and identifies areas where predictions are less accurate, allowing for future model modification.

.



*Fig.8.Bar graph of svr model*

**VII. CONCLUSION**

The application of Support Vector Regression (SVR) to predict traffic congestion is a prime example of the tremendous potential of machine learning in solving urban mobility challenges. The algorithm successfully forecasts traffic trends and congestion levels based on the analysis of historical traffic data, offering valuable insights to city planners and transportation agencies. Such predictions are instrumental in enhancing the flow of traffic, reducing journey delays, and making urban transportation systems more efficient as a whole.

The main objective of this research was to predict traffic density across different zones, detecting bottleneck areas and hours of heavy traffic. The model employs SVR to facilitate improved traffic management practices, optimize public transportation scheduling, and ensure smoother commute times. Additionally, the knowledge obtained from this predictive model can aid in the formulation of intelligent transportation.

**Future Scope of the Project:**

While the current model illustrates the potential of SVR, the following are various areas where it may be enhanced and developed:

1.Inclusion of real-time data: from sources like traffic sensors, cameras, and GPS-equipped vehicles enhances model responsiveness and accuracy. This would enable the system to give dynamic forecasts and updates, thereby optimizing real-time traffic management capability.

2.Smart Traffic Integration: The system can modify traffic light cycles, direct public transport, and regulate traffic flow based on anticipated rates of congestion.

3.Expanding its scope through the inclusion of traffic feeds from other cities makes the application that much broader and is an advanced tool for national or international traffic management. This can potentially increase collaboration between cities and encourage more insight into patterns of traffic.

4.Adding More Variables: The model here is based only on traffic information. Adding weather, accidents, or events will make the model even better and more complete forecast. Considering such outside factors would result in more accurate projections.

5.Deployment of Sophisticated Models: Deep learning and other machine learning methods can improve the accuracy of predictions, particularly in urban areas with chaotic or convoluted traffic patterns. The algorithms can analyze larger and more diverse datasets, providing more detailed insights into traffic trends.

In conclusion, while SVR is a reliable predictor of traffic flow, its future evolution with the integration of real-time data, broader coverage areas, and more parameters will further improve its predictive capability. These improvements can potentially transform urban traffic management, making cities more efficient, flexible, and sustainable.

**VIII .REFERENCES**

[1].Traffic prediction model using machine learning in intelligent transportation systems. (2023, August 10). IEEE Conference Publication | IEEE Xplore.

<https://ieeexplore.ieee.org/document/10245139>

[2]. Almukhalfi, H., Noor, A., & Noor, T. H. (2024). Traffic management approaches using machine learning and deep learning techniques: A survey. Engineering Applications of Artificial Intelligence, 133, 108147. <https://doi.org/10.1016/j.engappai.2024.108147>

[3]. Kim, Y., Tak, H., Kim, S., & Yeo, H. (2024). A hybrid approach of traffic simulation and machine learning techniques for enhancing real-time traffic prediction. Transportation Research Part C Emerging Technologies, 160, 104490. <https://doi.org/10.1016/j.trc.2024.104490>

[4]. Moreno, F. P., Rodríguez, F. I., Comendador, V. F. G., Jurado, R. D., Suárez, M. Z., & Valdés, R. M. A. (2024). Prediction of air traffic complexity through a dynamic complexity indicator and machine learning models. Journal of Air Transport Management, 119, 102632. <https://doi.org/10.1016/j.jairtraman.2024.102632>

[5]. Korkmaz, H., & Erturk, M. A. (2024). Prediction of the traffic incident duration using statistical and machine-learning methods: A systematic literature review. Technological Forecasting and Social Change, 207, 123621. <https://doi.org/10.1016/j.techfore.2024.123621>

[6]. Berhanu, Y., Schröder, D., Wodajo, B. T., & Alemayehu, E. (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for

safe routing on accident and Congestion-Prone road networks. Results in Engineering, 23, 102737. <https://doi.org/10.1016/j.rineng.2024.102737>

[7]. Comparative study of Machine learning Algorithms for traffic accident Prediction and Prevention. (2024,

June 24). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/10723916>

[8]. Machine Learning-Based Cellular traffic Prediction using data reduction techniques. (2024). IEEE Journals

&Magazine|IEEEXplore. <https://ieeexplore.ieee.org/document/10506912>

[9]. Machine learning based traffic prediction and congestion control algorithms in software defined networks. (2024, March 4). IEEE Conference Publication | IEEEXplore. <https://ieeexplore.ieee.org/document/10734570>

[10]. A vehicle speed prediction method integrating Multi-Source traffic information based on Informer. (2024, August 23). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/10703945>