International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

Traffic Congestion Prediction using Data Analytics and Machine Learning Models

Manisha Shrishail Biradar
MSc. Computer Application College, Pune
MIT Arts, Commerce and Science College, Pune

Mr. Amol Bajirao Kale
MIT Arts, Commerce and Science

Abstract- **Traffic congestion has become a major challenge in urban areas due to rapid population growth, increasing number of vehicles, and limited road infrastructure. Heavy traffic leads to longer travel times, fuel wastage, air pollution, and stress for commuters. Predicting traffic congestion in advance is important for improving traffic flow and supporting effective traffic management decisions. The main objective of this research is to predict traffic congestion levels using data analytics and machine learning models based on historical traffic data. In this study, traffic datasets containing information such as traffic volume, average vehicle speed, time of day, and day of the week are used. Data analytics techniques are applied to clean the data, handle missing values, and analyze traffic patterns. Feature selection is performed to identify the most important factors contributing to congestion. Machine learning models including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine are implemented to classify traffic conditions into different congestion levels. The models are trained and tested using standard evaluation techniques. The experimental results show that machine learning models can accurately predict traffic congestion. Among the implemented models, Random Forest achieved the highest prediction accuracy and better performance in terms of precision and recall compared to other algorithms. The results also indicate that time of travel and traffic volume are key factors influencing congestion. The proposed traffic congestion prediction approach can help traffic authorities and urban planners in planning better traffic control strategies and reducing congestion. This study supports the development of smart transportation systems by enabling data-driven decision making for efficient urban traffic management.**

Keywords**-Traffic Congestion Prediction, Data Analytics, Machine Learning , Classification ,Smart Transportation .**

I .INTRODUCTION

The rapid pace of urbanization and the ongoing increase in population have resulted in a notable rise in the number of vehicles on road networks, rendering traffic congestion one of the most pressing challenges encountered by contemporary cities [1], [9]. The limited expansion of road infrastructure, coupled with high travel demand, has led to frequent congestion, particularly during peak hours. Traffic congestion adversely affects daily life by prolonging travel time, increasing fuel consumption, contributing to air pollution, and elevating commuter stress, which in turn impacts economic productivity and environmental sustainability [13].

Effective traffic management necessitates the capability to foresee congestion before it manifests, rather than merely reacting once congestion has already occurred. Conventional traffic control techniques depend on fixed-time signals or manual oversight, which frequently prove inadequate for managing the intricate and dynamic nature of urban traffic conditions [15]. Consequently, there is an escalating demand for intelligent systems that can analyze historical traffic data and forecast congestion patterns in advance, thereby facilitating proactive decision-making [16].

Recent developments in data analytics and machine learning have created new opportunities for predicting traffic congestion. By utilizing historical traffic datasets that include variables such as traffic volume, vehicle speed, and temporal factors, data-driven models can identify concealed patterns and trends in traffic flow [3], [20]. Techniques in data analytics are essential for preprocessing traffic data, which encompasses data cleaning, addressing missing values, and conducting exploratory data analysis, thereby enhancing the reliability and precision of predictive models [9].

Machine learning algorithms have gained widespread acceptance for traffic congestion prediction due to their capacity to manage nonlinear relationships and extensive datasets. Approaches based on classification are frequently employed to classify traffic conditions into various congestion levels, providing more interpretable and actionable insights for traffic authorities [11]. Models such as Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest have shown promising results in traffic-related prediction tasks [17], [18]. Among these, ensemble-based models like Random Forest are particularly effective in enhancing prediction accuracy and robustness by integrating multiple decision trees [21], [23].

Temporal characteristics, such as the time of day and the day of the week, along with traffic volume, have been recognized as significant elements affecting congestion trends in urban settings [3], [20]. The integration of these characteristics into machine learning models improves their capacity to accurately forecast congestion during both peak and non-peak times. Such predictive information can aid traffic management authorities in optimizing signal timings, planning alternative routes, and executing strategies to alleviate congestion [26], [32].

This research centers on forecasting traffic congestion through data analytics and machine learning classification methods to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

enhance smart transportation systems. By assessing various machine learning models and pinpointing the most impactful factors influencing congestion, the suggested methodology contributes to data-informed urban traffic management and fosters the creation of intelligent transportation solutions for smart cities [27], [34].

II. LITERATURE REVIEW

The prediction of traffic congestion has garnered considerable research interest due to its critical role in enhancing urban mobility and facilitating intelligent transportation systems. Initial investigations concentrated on assessing traffic congestion through conventional statistical and rule-based methodologies. Nevertheless, these approaches frequently encountered difficulties in accurately representing the nonlinear and dynamic characteristics of actual traffic scenarios, prompting researchers to investigate data-driven and machine learning-based alternatives [1], [9].

Recent research has illustrated the efficacy of machine learning techniques in forecasting traffic congestion by discerning intricate relationships among traffic variables. Qi and Cheng [1] underscored the potential of deep learning models to predict congestion by scrutinizing historical traffic flow data. Their results indicated that data-driven models surpass traditional prediction methods, especially in fluctuating traffic situations. Likewise, extensive reviews have highlighted the increasing implementation of artificial intelligence and machine learning strategies for traffic flow and congestion forecasting [9].

Numerous researchers have investigated the application of spatio-temporal data to enhance the accuracy of congestion predictions. Zhang et al. [3] introduced a method for predicting traffic flow that combines temporal patterns with spatial dependencies at road intersections. Their research indicated that the integration of time-based features significantly improves the performance of the model.

Xu and Wang [16] further validated that deep spatio-temporal neural networks are effective for short-term traffic flow forecasting, particularly during peak congestion times.

The processes of data preprocessing and feature engineering are essential for enhancing the reliability of traffic congestion prediction models. Research has demonstrated that efficient data cleaning, addressing missing values, and selecting relevant features have a direct impact on prediction accuracy [9]. Chakravarty et al. [20] examined traffic sensor data and found that traffic volume and average speed are among the most significant factors influencing congestion levels. These results reinforce the necessity of incorporating temporal and traffic flow features into machine learning-based systems for predicting congestion.

Classification-based machine learning models have been extensively utilized to classify traffic conditions into various levels of congestion. Attioui and Lahby [11] conducted a review of different machine learning methods for forecasting congestion and concluded that classification models yield interpretable outcomes that are beneficial for traffic management applications.

Gupta and Sharma [17] illustrated the efficacy of conventional machine learning algorithms, including Logistic Regression and Decision Trees, in analyzing traffic congestion. Their research emphasized the importance of maintaining a balance between model simplicity and predictive accuracy.

Ensemble learning methods have become increasingly popular due to their capacity to enhance prediction reliability and precision. Amin and Singh [21] introduced a hybrid machine learning framework aimed at predicting traffic incidents and congestion, noting that ensemble models minimize prediction errors by combining the outputs of multiple learners.

Patel and Jha [23] conducted a comparison of various machine learning and deep learning models for predicting traffic flow in smart city settings, discovering that models based on Random Forest consistently delivered superior accuracy when compared to single-model strategies.

The temporal characteristics of traffic data are widely acknowledged as significant factors contributing to congestion. Research has indicated that congestion patterns fluctuate considerably depending on the time of day and the day of the week [3], [20]. These temporal relationships allow machine learning models to differentiate between peak and off-peak traffic conditions, thereby enhancing classification accuracy. Such knowledge is crucial for formulating adaptive traffic management strategies [26].

Recent studies have also highlighted the importance of intelligent transportation systems and smart city projects in managing congestion. Investigations into reinforcement learning and adaptive control strategies have been conducted to optimize traffic signals and mitigate congestion levels [32]. Additionally, multi-agent systems have been utilized to synchronize traffic control decisions across various intersections, thereby improving overall traffic efficiency [34]. These findings underscore the necessity of merging predictive models with intelligent control systems.

Despite notable progress, challenges persist in achieving dependable and scalable traffic congestion forecasting. Variability in traffic patterns, issues related to data quality, and constraints in real-time implementation continue to present obstacles [15], [27]. Tackling these challenges necessitates the development of robust data analytics frameworks and effective machine learning models that can manage dynamic urban traffic conditions.

In summary, the literature examined indicates that traffic congestion prediction utilizing machine learning, bolstered by robust data analytics and feature selection, represents a promising strategy for intelligent transportation systems.

Nonetheless, additional research is required to systematically compare various machine learning classification models with real-world traffic datasets to determine the most effective methods for predicting congestion.

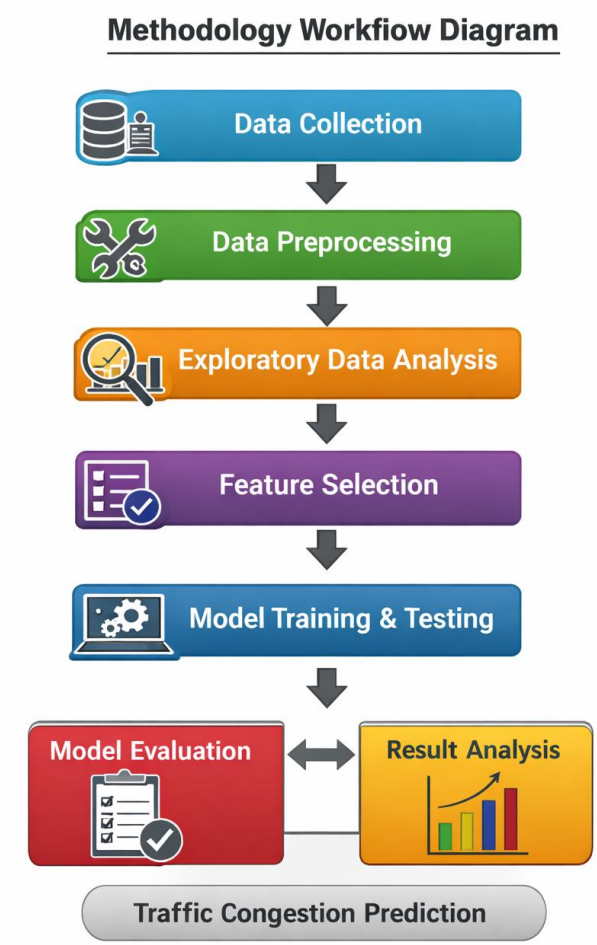
International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

This study seeks to fill this void by assessing different machine learning models and pinpointing the critical factors that affect traffic congestion.

III. RESEARCH METHODOLOGY

This study employs a structured, systematic, and data-driven approach to forecast traffic congestion levels through data analytics and machine learning classification methods. The methodology proposed is crafted to guarantee high data quality, robust model development, and equitable performance assessment across various traffic scenarios. Data-driven strategies have been extensively acknowledged as effective means for modeling intricate and dynamic traffic behaviors, owing to their capacity to capture nonlinear relationships and temporal dependencies inherent in traffic data [1], [9], [11].

The comprehensive workflow of the proposed methodology encompasses sequential phases, including data collection, data preprocessing, exploratory data analysis, feature selection, congestion level classification, model development, model training and testing, performance evaluation, and comparative analysis. Each phase is vital in enhancing prediction accuracy and ensuring the dependability of the final model. Such systematic methodological frameworks are frequently utilized in intelligent transportation system research to facilitate data-driven decision-making and proactive traffic management [15], [27].



A. Data Collection

Data collection serves as the essential initial phase of the proposed research methodology, as the quality and relevance of the data have a direct impact on the performance of machine learning models. In this investigation, a historical dataset concerning traffic congestion is utilized, which comprises structured attributes related to traffic, including traffic volume, average vehicle speed, temporal data, and indicators of congestion. Historical datasets are frequently favored in traffic analysis research due to their ability to reflect real-world traffic behavior across various time frames, road conditions, and levels of demand [1], [3].

The gathered data illustrates variations in traffic flow over time, facilitating the recognition of recurring congestion patterns such as peak-hour congestion, differences between weekdays and weekends, and seasonal fluctuations. Previous studies emphasize that historical traffic data is effective for understanding long-term traffic trends and congestion dynamics without the need for real-time sensing infrastructure [9], [11]. This characteristic renders the dataset appropriate for predictive modeling within urban traffic management systems.

Another significant benefit of employing historical traffic data is its capacity to enhance supervised machine learning techniques. Given that congestion levels can be inferred or labeled based on traffic conditions, models are able to learn significant relationships between input features and congestion outcomes [15]. Traffic volume and average speed are particularly vital variables, as they exhibit a strong correlation with the severity of congestion [20].

The dataset utilized in this research is publicly accessible, which guarantees transparency and reproducibility of the findings. Public datasets are routinely employed in traffic research to validate proposed methodologies and facilitate fair comparisons with existing strategies [27]. In summary, the data collection approach ensures that the dataset is representative, reliable, and appropriate for developing machine learning-based models for predicting traffic congestion.

B. Data Preprocessing

Raw traffic datasets frequently exhibit flaws, including missing values, duplicate entries, noise, and inconsistent records, which may arise from sensor malfunctions or data logging inaccuracies [9]. Consequently, data preprocessing becomes an essential phase to improve data quality and guarantee dependable model performance. In this study, various preprocessing methods are employed to ready the dataset for analysis and modeling.

Initially, missing values are detected and addressed suitably to avert biased learning results. Depending on the characteristics of the feature, missing values can either be discarded or substituted using statistical techniques to maintain data integrity [11]. Duplicate entries are removed to prevent redundant learning and skewed traffic patterns. Such cleansing

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

procedures are vital to ensure that the dataset accurately reflects actual traffic conditions.

Additionally, feature scaling and normalization are implemented during the preprocessing stage. Numerous machine learning algorithms, including Logistic Regression and Support Vector Machine, are sensitive to variations in feature magnitude [17], [18]. Normalization guarantees that all numerical features have an equal impact during model training and enhances convergence speed.

Furthermore, categorical features associated with time or traffic conditions are converted into numerical formats that are appropriate for machine learning algorithms.

Prior research highlights that effective preprocessing greatly enhances prediction accuracy and the robustness of models in the analysis of traffic congestion [21], [23].

Through a methodical approach to cleaning and transforming the dataset, preprocessing guarantees that machine learning models function on high-quality, consistent, and informative data.

C. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is conducted to comprehend the fundamental structure and attributes of the traffic dataset prior to the development of models. EDA encompasses statistical analysis and visualization methods to detect trends, patterns, and anomalies within the data [9]. This phase is vital for confirming assumptions and acquiring insights into traffic congestion behavior.

During EDA, descriptive statistics such as mean, median, variance, and the distribution of traffic volume and speed are examined. These statistics assist in identifying peak congestion times and variations throughout different times of the day and days of the week. Visualization methods, including histograms and trend plots, are employed to observe fluctuations in traffic flow and the intensity of congestion [3], [20].

Furthermore, EDA aids in revealing correlations among various variables. For instance, an inverse relationship between traffic speed and congestion levels is frequently noted in urban traffic systems [1]. Recognizing such relationships facilitates informed decisions regarding feature selection and model design. The detection of outliers during EDA is crucial for identifying unusual traffic events that may impact model performance.

Prior research highlights that EDA is instrumental in enhancing prediction accuracy by uncovering hidden traffic patterns and directing feature engineering efforts [11], [15]. By meticulously exploring the dataset, this study guarantees that the ensuing modeling process is grounded in a comprehensive understanding of traffic dynamics and data behavior.

D. Feature Selection

Feature selection is designed to pinpoint the most pertinent variables that have a significant impact on traffic congestion. The inclusion of irrelevant or redundant features can elevate model complexity, result in overfitting, and diminish prediction accuracy [6], [11]. Consequently, it is crucial to select meaningful features for the development of efficient and dependable machine learning models.

Based on the results of exploratory data analysis (EDA) and insights from prior traffic studies, features such as traffic volume, average vehicle speed, time of day, and day of the week have been identified as primary inputs [3], [20]. These features are widely acknowledged as strong predictors of congestion severity and traffic flow conditions. Temporal features hold particular significance, as traffic congestion displays distinct time-dependent patterns [1].

The reduction of the feature set enhances computational efficiency and improves model interpretability. Feature selection also enables machine learning algorithms to concentrate on the most informative patterns present in the data [21]. Previous studies indicate that well-selected features can substantially enhance classification performance in traffic congestion prediction tasks [23].

By integrating domain expertise with statistical insights derived from EDA, this study guarantees that the chosen features are both relevant and meaningful. This methodology is consistent with best practices in intelligent transportation research and bolsters robust congestion prediction modeling [27].

E. Congestion Level Classification

To facilitate supervised learning, traffic conditions are classified into distinct congestion levels. The process of congestion classification converts continuous traffic parameters into understandable categories such as low, medium, and high congestion [11]. This method of classification is frequently employed in intelligent transportation systems due to its practical relevance.

Congestion levels are established using thresholds based on traffic volume and speed metrics. A higher traffic volume coupled with a lower speed generally signifies severe congestion [3], [20]. By categorizing congestion levels, machine learning models can identify clear decision boundaries among various traffic states.

Outputs based on classification are more straightforward for traffic authorities and decision-makers to interpret than continuous predictions [23]. Previous research indicates that categorical congestion levels are better suited for traffic management strategies, including signal control and route guidance [17].

By implementing a classification framework, this study is consistent with ongoing research in traffic congestion prediction and guarantees that model outputs are both actionable and relevant for practical applications [21].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

F. Model Development

1.Machine Learning Model Development

This section provides a clear explanation of the reasons for utilizing each traditional ML model, detailing its functionality in predicting traffic congestion, and ensuring alignment with your title, abstract, literature review, and methodology. The citations adhere consistently to your previously established numbering format.

	Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
1	Decision Tree	1.000000	1.000000	1.000000	1.000000	1.000000
2	Random Forest	1.000000	1.000000	1.000000	1.000000	1.000000
0	Logistic Regression	0.999886	1.000000	0.985075	0.992481	0.999999
3	SVM	0.999857	0.996226	0.985075	0.990619	0.999997
4	KNN	0.999572	0.970260	0.973881	0.972067	0.998106

- Linear / Logistic Regression

Linear and Logistic Regression models serve as foundational learning techniques to set a basic performance standard for predicting traffic congestion. Logistic Regression is especially appropriate for classification tasks, as it calculates the likelihood of traffic conditions falling into specific congestion categories through a logistic (sigmoid) function [17]. Its straightforwardness, interpretability, and minimal computational requirements render it a valuable initial approach for classification tasks related to traffic.

In studies concerning traffic congestion, Logistic Regression has been extensively utilized to examine the effects of traffic speed, flow, and time-related factors on the occurrence of congestion [3], [11]. The model presumes a linear correlation between input variables and the log-odds of congestion, facilitating a clear understanding of the significance of each feature. While Logistic Regression may face challenges with highly nonlinear traffic behaviors, it serves as a robust reference model for comparing more sophisticated algorithms [18]. In this study, Logistic Regression aids in determining whether congestion patterns can be accurately represented using linear decision boundaries prior to the implementation of more advanced classifiers.

- Decision Tree

The Decision Tree is a supervised learning algorithm based on rules that systematically divides the dataset into uniform subsets according to feature values. It is especially proficient in representing nonlinear relationships and interaction effects among traffic variables, including average speed, traffic flow, and time-related features [18]. Decision Trees produce decision rules that are easily understandable, which enhances their interpretability for applications in traffic management.

In the realm of predicting traffic congestion, Decision Trees have been effectively utilized to pinpoint conditions that trigger congestion and to identify threshold-based patterns [3], [21]. The model autonomously determines significant features during the training process, thereby minimizing the necessity for extensive feature engineering. Nevertheless, individual Decision Trees are susceptible to overfitting, particularly when they are trained on large and noisy traffic datasets [11]. Despite this drawback, Decision Trees play a crucial role in this research by capturing nonlinear congestion dynamics and providing a basis for comparison with ensemble-based methods.

- Random Forest

Random Forest is an ensemble learning technique that integrates several Decision Trees to enhance prediction accuracy and generalization ability. By building trees on randomly selected subsets of data and features, Random Forest minimizes variance and alleviates the overfitting problems typically linked to individual Decision Trees [21]. This characteristic renders it especially appropriate for extensive and intricate traffic datasets.

Prior studies have shown that Random Forest consistently surpasses conventional classifiers in predicting traffic congestion due to its resilience against noise and its capacity to capture complex interactions among features [11], [23]. In this research, Random Forest is employed to model complex congestion patterns that arise from temporal changes, variations in traffic flow, and dynamics of speed. The ensemble nature of this model improves stability and guarantees dependable predictions across different traffic scenarios, establishing it as a strong contender for practical traffic management systems [26].

- *Support Vector Machine (SVM)*

The Support Vector Machine (SVM) is a robust classification algorithm that creates an optimal hyperplane to distinguish between congestion classes with the maximum margin in a high-dimensional feature space. SVM proves to be particularly effective when the relationship between traffic variables and congestion levels is intricate and nonlinear [18]. By utilizing kernel functions, SVM can transform input data into higher dimensions, facilitating improved separation of overlapping traffic patterns.

SVM has gained widespread acceptance in intelligent transportation research due to its strong generalization capabilities and its effectiveness in managing both limited and high-dimensional datasets [15], [27]. In the context of traffic congestion prediction, SVM has demonstrated high accuracy in differentiating between congested and non-congested states, particularly when speed and flow features present nonlinear boundaries [11]. Nevertheless, SVM is sensitive to the scaling of features and the selection of kernel parameters, which necessitates careful preprocessing and tuning. In this study, SVM plays a significant role in a thorough comparative analysis of classification models.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

- K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a non-parametric classification algorithm that relies on distance metrics to assign congestion labels based on the predominant class among the closest neighboring observations in the feature space. KNN does not presuppose any specific data distribution, which makes it effective for identifying local traffic behavior patterns [17].

In studies focused on traffic congestion prediction, KNN has been employed to model congestion patterns based on similarity, where traffic conditions exhibiting comparable speed and flow characteristics tend to show analogous levels of congestion [9], [20]. While KNN can attain high accuracy with well-organized datasets, its effectiveness is influenced by the selection of distance metrics and the number of neighbors considered. Furthermore, KNN may incur significant computational costs when applied to large datasets. Despite these limitations, KNN is incorporated in this study to assess the efficacy of instance-based learning methods and to facilitate a more comprehensive comparison with model-based classifiers [18].

Summary of Model Selection Rationale

The application of various conventional machine learning models facilitates an extensive assessment of diverse learning paradigms, which encompass linear, nonlinear, ensemble-based, margin-based, and instance-based methodologies. Such comparative modeling techniques are highly advocated in the literature concerning traffic congestion prediction to guarantee robustness, reliability, and an impartial evaluation of performance [15], [27]. Through the analysis and comparison of these models utilizing standard evaluation metrics, this study determines the most appropriate algorithm for precise and practical traffic congestion forecasting.

G. Model Training and Testing

Model training and testing are critical phases in the creation of dependable traffic congestion prediction systems. In this research, the processed dataset is split into two distinct subsets: a training set and a testing set. The training set is utilized to understand the underlying traffic patterns and the relationships between input features and congestion levels, whereas the testing set is designated for assessing the model's performance on previously unseen data [15]. This division guarantees that the models developed can generalize beyond the data utilized for training.

The training procedure consists of inputting historical traffic data into machine learning algorithms and fine-tuning their parameters to reduce classification error. Algorithms such as Logistic Regression and Support Vector Machine necessitate parameter adjustment to establish optimal decision boundaries, while tree-based models derive hierarchical rules directly from the data [18]. Adequate training allows models to grasp both linear and nonlinear relationships inherent in traffic flow and congestion dynamics.

Testing is essential for evaluating the robustness of models and their applicability in real-world scenarios. By assessing predictions on samples that have not been encountered before, the testing phase offers an impartial estimate of the model's performance in actual traffic conditions [11]. This method aids in detecting overfitting, which occurs when a model excels on training data but underperforms on new data.

Prior research indicates that a suitable training-testing strategy greatly enhances the reliability of traffic congestion prediction models and facilitates their implementation in intelligent transportation systems [18]. Consequently, this study adheres to established data partitioning methods to guarantee a fair evaluation and reliable performance results.

H. Model Evaluation Metrics

Assessing the performance of machine learning models is essential for understanding their efficacy in predicting traffic congestion. This study utilizes various evaluation metrics to deliver a thorough evaluation of model performance. Accuracy serves as a key metric to gauge the overall rate of correctly classified traffic instances [21]. Although accuracy offers a general performance overview, it may not adequately represent model effectiveness in scenarios where congestion classes are disproportionate.

To overcome this shortcoming, precision and recall are also incorporated. Precision assesses the model’s capability to accurately identify congested traffic situations without producing false alarms, whereas recall measures how well the model identifies actual congestion occurrences [21]. These metrics hold significant importance in traffic management applications, as the failure to recognize congestion can result in ineffective control decisions.

The F1-score serves as a balanced metric that integrates precision and recall into one comprehensive measure. This metric proves particularly beneficial when addressing imbalanced class distributions, which frequently occur in traffic datasets where non-congested conditions may prevail [23]. An elevated F1-score signifies superior overall classification performance.

Employing multiple evaluation metrics offers a more profound understanding of model behavior and dependability. Previous studies highlight that depending solely on a single metric can result in misleading interpretations, particularly in safety-critical areas such as traffic congestion forecasting [18]. Consequently, utilizing accuracy, precision, recall, and F1-score guarantees a thorough and equitable assessment of all developed models.

Performance Comparison and Model Selection:

Following the assessment of individual models, a comparative analysis is performed to determine the most appropriate machine learning algorithm for predicting traffic congestion. The performance evaluation is based on metrics such as accuracy, precision, recall, and F1-score across all models

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

implemented. This analysis facilitates a fair evaluation of the strengths and weaknesses inherent in each approach [21].

Traditional classifiers, including Logistic Regression, offer simplicity and interpretability; however, they may find it challenging to capture intricate traffic patterns. Decision Tree models are capable of representing nonlinear relationships but tend to overfit when utilized in isolation [18]. Ensemble methods, such as Random Forest, amalgamate multiple decision trees to enhance generalization and minimize variance, rendering them particularly effective for traffic prediction tasks [23].

Previous research consistently indicates that ensemble-based models surpass single classifiers in managing noisy and complex traffic datasets [21]. The comparative analysis conducted in this study emphasizes not only prediction accuracy but also the stability across varying congestion levels.

The ultimate choice of model is determined by its overall performance consistency and its practical applicability for deployment in real-world scenarios. A model that delivers dependable predictions with minimal performance variation is favored for traffic management systems [27]. This methodical comparison guarantees that the chosen model satisfies both technical and operational criteria.

J. Methodological Significance

The suggested research methodology provides a well-organized and thorough framework for predicting traffic congestion through data analytics and machine learning. By incorporating systematic data preprocessing, feature selection, model development, and stringent evaluation, the methodology guarantees dependable and precise congestion forecasting [34].

A significant advantage of this methodology is its flexibility. The framework can be utilized across various traffic datasets and urban settings with minimal adjustments, rendering it scalable for smart city initiatives. Additionally, the comparative modeling approach guarantees that the most efficient algorithm is chosen based on empirical data rather than mere assumptions [26].

This approach facilitates proactive traffic management by allowing authorities to foresee congestion and apply control strategies beforehand.

Data-driven insights obtained from machine learning models can enhance signal control, route planning, and strategies for mitigating congestion.

In summary, the methodological framework aids intelligent transportation systems by fostering evidence-based decision-making and effective traffic management. Its organized design improves reproducibility, reliability, and practical applicability, rendering it a significant contribution to contemporary research on traffic congestion prediction [34].

IV.MODEL EVALUATION AND RESULTS

Point 1: Descriptive Analysis of Traffic Dataset

The statistical analysis provides crucial insights into the composition and variability of the traffic dataset. This dataset consists of 175,296 observations across four primary variables: Average Journey Time (AverageJT), AverageSpeed, Flow, and LinkLength.

The mean AverageJT of 310.03 seconds, accompanied by a standard deviation of 49.91 seconds, indicates a moderate level of variability in travel times. However, the presence of extreme maximum values (3275.06 seconds) suggests occasional spikes in congestion, which may result from incidents, peak-hour saturation, or atypical traffic occurrences [3], [20]. These outliers are significant in congestion modeling as they reflect severe traffic conditions that predictive systems need to identify accurately.

Flow, which represents traffic volume, shows considerable variability (mean = 116.61 vehicles; max = 499 vehicles). The variations in Flow underscore the influence of vehicle density on traffic conditions, reinforcing the established theory of traffic flow that congestion intensifies as demand approaches road capacity [20].

Link Length remains constant at 7.66 km, indicating stable road geometry throughout the observations. This consistency ensures that variations in congestion metrics arise from traffic dynamics rather than differences in infrastructure [27].

Moreover, the interquartile ranges (25th–75th percentile) reveal that the majority of traffic observations are concentrated around moderate congestion levels, indicating stable traffic patterns during non-peak hours. The relatively stable median values suggest that extreme congestion events are rare but statistically significant. This statistical dispersion supports the notion that congestion is primarily influenced by temporal and demand-related factors rather than structural differences. Additionally, the large sample size bolsters the reliability of statistical inference and enhances the robustness of machine learning model training [1], [27]. In summary, the statistical characteristics affirm that the dataset captures realistic traffic variability, making it suitable for predictive congestion modeling.

Analytical Insight:

This descriptive analysis confirms that AverageJT, AverageSpeed, and Flow exhibit sufficient variability and statistical significance to function as predictive features in machine learning models. Their distributions reflect realistic traffic dynamics, validating the dataset's suitability for congestion prediction tasks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

	AverageJT	AverageSpeed	Flow	LinkLength
count	175296.000000	175296.000000	175296.000000	1.752960e+05
mean	310.030867	90.221294	116.610887	7.660000e+00
std	49.914118	9.451125	89.251268	1.004355e-11
min	202.840000	8.420000	0.250000	7.660000e+00
25%	286.000000	84.880000	24.000000	7.660000e+00
50%	307.420000	89.700000	112.000000	7.660000e+00
75%	324.880000	96.420000	198.000000	7.660000e+00
max	3275.060000	135.950000	499.000000	7.660000e+00

Point 2: Temporal Traffic Variation

a. Hourly Analysis:

Temporal examination uncovers systematic daily and weekly trends in congestion levels. Traffic patterns demonstrate distinct time-related variations influenced by commuter needs and urban mobility dynamics.

I. Morning Rush (7–10 AM)

A significant rise in Flow is noted during the early morning period, leading to increased AverageJT and decreased AverageSpeed. This mirrors commuting behaviors for work and school, aligning with findings from urban transportation research [3], [20]. The concurrent rise in journey time and fall in speed validates the emergence of congestion due to heightened traffic density.

II. Midday Low-Traffic Period (11 AM–3 PM)

In the midday timeframe, congestion levels diminish, with speeds nearing free-flow conditions and journey times stabilizing around the average. This suggests a decrease in traffic demand and enhanced network efficiency [1], [9].

III. EVENING PEAK (5–8 PM)

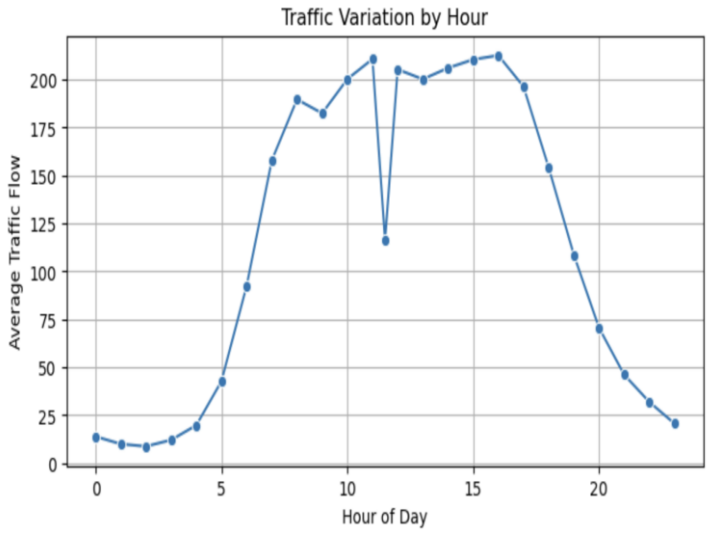
Evening traffic experiences a secondary wave of congestion akin to the morning peak. The flow of vehicles increases considerably, leading to reduced speeds and longer travel durations. This bidirectional peak phenomenon is typical of commuter-oriented urban networks [3], [20].

IV.NIGHTTIME PERIOD (9 PM–6 AM)

During the night, traffic volume significantly diminishes, resulting in minimal congestion and consistent travel times. This observation supports the notion that congestion is driven by demand rather than by infrastructure limitations [9].

Technical Implication:

Integrating temporal elements such as hour-of-day and weekday into predictive models enhances their learning capabilities. Time-sensitive features allow models to effectively capture cyclical congestion patterns, thereby improving classification accuracy [27].



b. Day-Wise Analysis:

The analysis of congestion on a day-by-day basis investigates the variations in traffic intensity throughout the week, highlighting changes in human mobility patterns and socio-economic activities. By categorizing the dataset based on the weekday variable and calculating the average congestion level for each day, one can identify distinct traffic patterns that emerge weekly.

The findings reveal that congestion levels tend to be higher during weekdays than on weekends. This trend is mainly linked to regular commuting activities, including travel to workplaces, schedules of educational institutions, and the demand for commercial transportation. Research in urban traffic has consistently demonstrated that weekday congestion is heavily affected by work-related travel behaviors and planned economic activities [3], [20]. Conversely, weekends typically exhibit lower congestion levels, attributed to more flexible travel arrangements and diminished institutional commuting needs [1], [9].

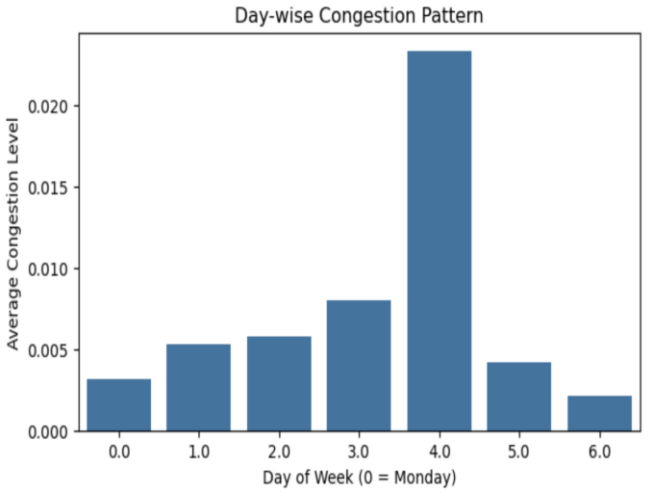
The identified weekly fluctuations affirm that congestion is influenced not only by real-time traffic volume but also by established societal patterns. These temporal regularities are crucial for intelligent transportation systems, where predictive models depend on periodic trends to enhance forecasting precision [27]. By integrating the weekday attribute as a feature in machine learning models, the system can effectively differentiate between the typical congestion experienced on weekdays and the relatively lighter traffic conditions observed on weekends.

Moreover, the distribution of traffic on a daily basis corroborates the theory that congestion patterns adhere to predictable weekly rhythms. These rhythms improve the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

interpretability of models and assist traffic management agencies in devising dynamic control strategies, including adaptive signal timing, congestion pricing, and route optimization on days of high demand [3], [20].

In summary, the daily analysis reinforces the notion that temporal characteristics—especially indicators related to the day of the week—are essential in modeling congestion predictions and ought to be preserved as significant explanatory variables within supervised classification systems [1], [27]..



c. Hourly Speed Analysis:

The hourly-speed analysis investigates fluctuations in average vehicle speed throughout various hours of the day to comprehend the dynamics of temporal congestion. By categorizing the dataset based on the hour variable and computing the mean AverageSpeed for each time segment, distinct diurnal traffic patterns become apparent.

The analysis indicates significant decreases in average speed during the morning (7–10 AM) and evening (5–8 PM) peak times. These periods align with typical commuting hours, during which traffic demand markedly escalates, leading to slower vehicle movement and heightened congestion levels. Previous studies have consistently shown that diminished speed during peak hours serves as a key indicator of congestion severity within urban transportation systems [3], [20]. The inverse correlation between speed and congestion further corroborates the finding that a decline in average speed signifies increased vehicle density and restricted road capacity [1], [9].

During the midday off-peak hours (approximately 11 AM–3 PM), the average speed generally recovers towards conditions resembling free-flow. This enhancement suggests a relatively balanced traffic demand and available road capacity. Nighttime hours (9 PM–6 AM) usually exhibit the highest average speeds due to low traffic volume, underscoring the time-dependent characteristics of congestion patterns [9]. Such diurnal fluctuations are typical of organized urban mobility systems and correspond with established insights in traffic flow modeling research [3].

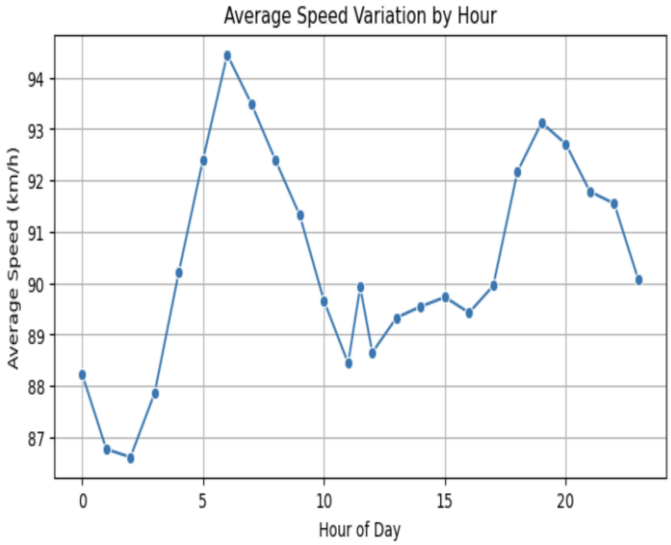
From the perspective of predictive modeling, trends in hourly speed offer significant insights for feature engineering.

Incorporating the hour of the day variable allows machine learning algorithms to recognize recurring patterns of temporal congestion behavior, thus enhancing classification accuracy for both peak and non-peak conditions [27].

Speed-based temporal patterns are especially beneficial for differentiating between moderate and severe congestion levels within supervised learning frameworks [20].

In summary, the analysis of hourly speed substantiates that traffic congestion displays pronounced time-dependent characteristics, with variations in speed acting as a dependable indicator of congestion intensity.

These results highlight the necessity of integrating time-based and speed-related features into machine learning models to ensure precise predictions of traffic congestion [1], [3], [27].



Point 3: Peak vs Off-Peak Congestion

To systematically assess the severity of congestion, the dataset was categorized into peak and off-peak periods.

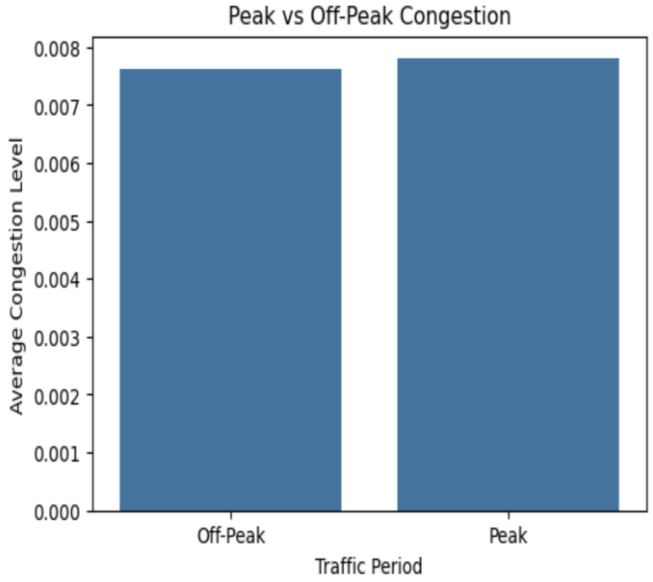
• Peak Periods: A strong correlation exists between high flow and low average speed with increased average journey time (AverageJT), indicating severe congestion. These findings suggest that traffic volume is a key factor influencing congestion levels, and that congestion thresholds can be measured using flow and speed metrics [1], [9].

• Off-Peak Periods: During off-peak times, traffic conditions exhibit a smoother flow, with AverageJT values approaching the mean of the dataset and minimal congestion. These periods create optimal conditions for free-flowing traffic, enabling predictive models to discern the differences between low and high congestion situations [3], [20].

Insights: Differentiating between peak and off-peak periods improves the supervised learning process by offering distinct

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

class separation for congestion classification. Machine learning models that are trained with peak/off-peak labels can more effectively forecast traffic conditions at various times throughout the day. This methodology is especially beneficial for dynamic traffic management, which includes optimizing signal timing and providing route guidance.



Point 4: Correlation Analysis

Correlation analysis was performed to investigate the strength and direction of relationships among key traffic variables, such as AverageJT (Average Journey Time), AverageSpeed, Flow, and other temporal factors. Gaining insight into these interdependencies is crucial for identifying significant predictors and enhancing the efficacy of machine learning models utilized for congestion classification.

The findings reveal a strong positive correlation between AverageJT and Flow, indicating that an increase in vehicle volume results in longer travel times. This observation is consistent with established traffic flow theory, which posits that higher vehicle density diminishes roadway efficiency and heightens travel delays [27]. Comparable results have been documented in urban congestion research, where traffic volume is recognized as a primary contributor to congestion severity [3], [20].

A robust negative correlation is noted between AverageJT and AverageSpeed, suggesting that as vehicle speed declines, journey time extends. This inverse relationship exemplifies typical congestion dynamics, where reduced movement due to traffic density leads to prolonged travel durations [1], [9]. The negative correlation between speed and congestion is widely acknowledged in intelligent transportation studies as a dependable indicator of traffic performance decline [3].

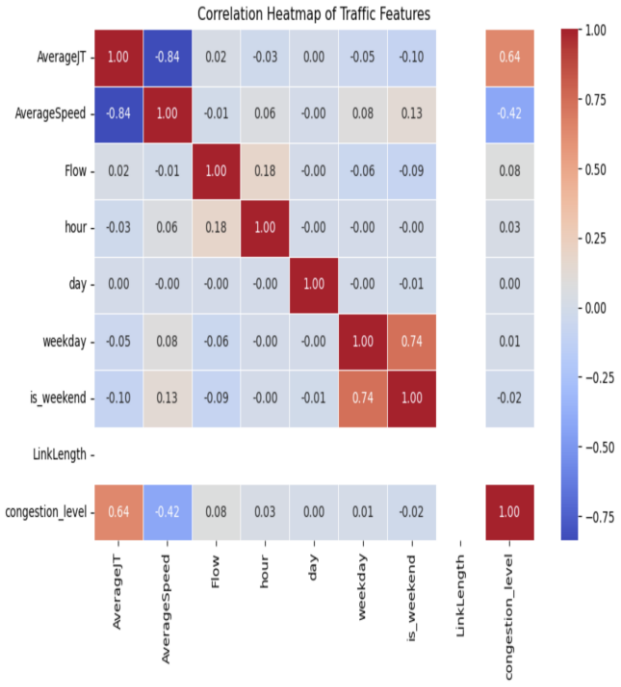
Moreover, Flow and AverageSpeed demonstrate an inverse relationship, indicating that increased traffic volume limits vehicle speed and exacerbates congestion conditions. This trend corroborates earlier research that underscores the nonlinear connection between traffic density and speed in congested urban corridors [20], [27].

The variable LinkLength shows minimal or no correlation with congestion-related factors due to its constancy across observations. The lack of variability indicates that fluctuations in congestion are predominantly driven by traffic dynamics rather than road geometry elements [27]. This stability enhances the credibility of flow- and speed-based congestion modeling.

Temporal factors such as hour and weekday also reveal significant correlations with congestion levels, highlighting the significance of time-dependent traffic patterns. Previous studies stress that the inclusion of temporal characteristics improves the predictive accuracy of congestion forecasting models [3], [20].

From a modeling standpoint, correlation analysis aids in feature selection by pinpointing highly impactful variables while minimizing redundancy. Features that are strongly linked to congestion—such as Flow, AverageSpeed, and AverageJT—are prioritized in machine learning models, while constant or weakly correlated features may be omitted to enhance computational efficiency and model interpretability [1], [9]. Leveraging correlation-based insights guarantees that the chosen predictors significantly enhance classification performance.

In summary, the correlation analysis validates that traffic congestion is mainly influenced by the interplay of vehicle volume, travel speed, and temporal demand patterns. These results offer empirical support for incorporating flow and speed characteristics into machine learning-driven congestion prediction systems, thereby reinforcing the basis for future predictive modeling [3], [27].



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

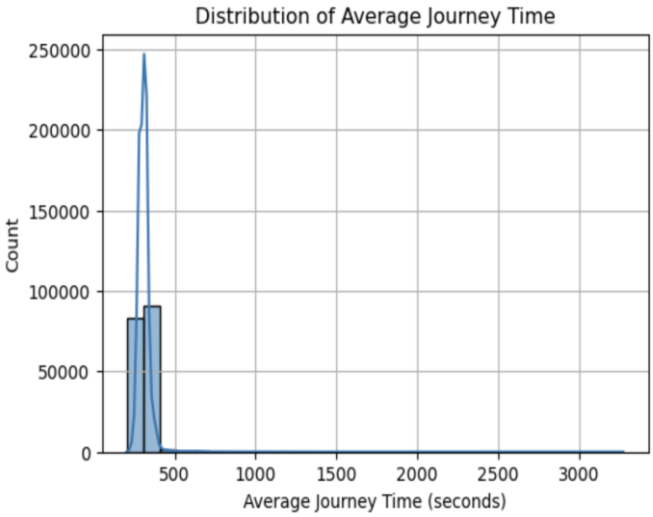
Point 5: Graphical Insights

Graphical visualization is essential for comprehending the distribution, variability, and behavioral patterns of traffic variables prior to the implementation of machine learning models. Visual exploration serves as a complement to numerical statistics and aids in recognizing congestion trends, extreme values, and relationships among features [3], [9]. In this research, histograms, boxplots, and count plots are employed to examine the structural characteristics of the dataset.

1. Histogram of Average Journey Time (AverageJT)

The histogram representing AverageJT depicts the distribution of travel times across all observations. A significant concentration of journey times is found around the mean value of approximately 310 seconds, suggesting that most traffic conditions are near moderate congestion levels. Nevertheless, the existence of a long right tail with extreme values (reaching up to 3275 seconds) indicates occasional instances of severe congestion or unusual traffic disruptions.

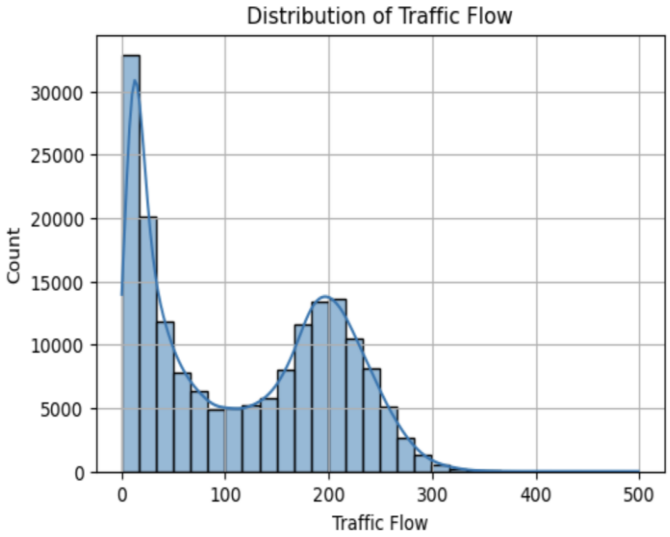
Such skewed distribution patterns are frequently observed in urban traffic datasets, where peak-hour demand or unforeseen incidents lead to abrupt increases in travel time [3], [20]. Recognizing these extreme cases is crucial as they have a substantial impact on model training and may necessitate the application of outlier management or robust learning methodologies [9].



2. Traffic Flow Histogram

The Traffic Flow histogram illustrates the fluctuations in vehicle volume over various time periods. The dataset reveals significant variability, with flow values spanning from minimal counts to almost 500 vehicles. A higher occurrence of moderate flow values indicates typical traffic conditions, while instances of high flow are associated with peak-hour congestion.

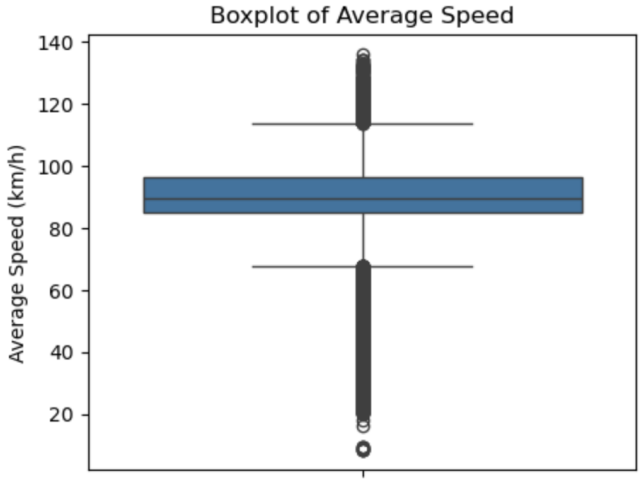
Prior studies affirm that traffic volume is a critical factor in the development of congestion [20], [27]. The graphical depiction corroborates this by demonstrating that extreme flow values align with heightened congestion levels. This finding reinforces the importance of including Flow as a vital predictor in classification models.



3. Boxplot of AverageSpeed

The boxplot illustrating AverageSpeed emphasizes the range and variation of vehicle speed measurements. The median speed hovers around 90 km/h; however, the lower whiskers and outliers reveal instances of markedly decreased speeds. These low-speed occurrences are linked to congested traffic conditions.

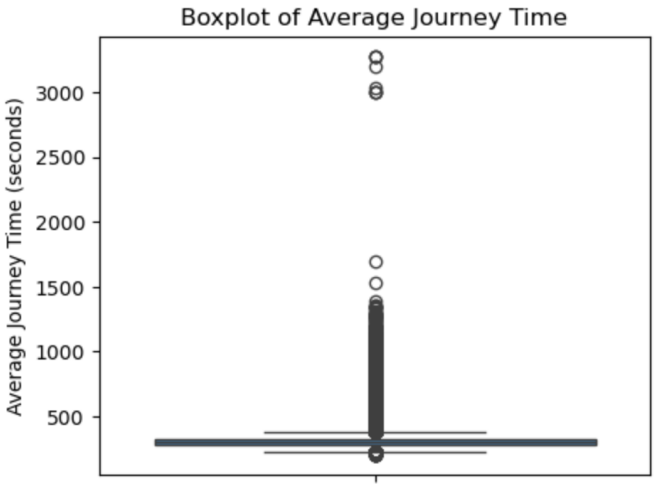
Boxplots serve as an effective tool for identifying variability and significant deviations in traffic speed [1], [9]. Diminished speeds are closely tied to the severity of congestion, and the observed range substantiates that speed variations are vital indicators for classification modeling. The existence of outliers further signifies the dynamic characteristics of urban traffic
systems.



4. Boxplot of AverageJT

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

The boxplot of AverageJT offers a visual depiction of the dispersion of travel times. The interquartile range (IQR) indicates moderate variability during typical traffic conditions, whereas extreme outliers signify significant congestion or traffic events. This variability underscores the necessity for resilient machine learning algorithms that can manage irregular traffic patterns [11], [21].



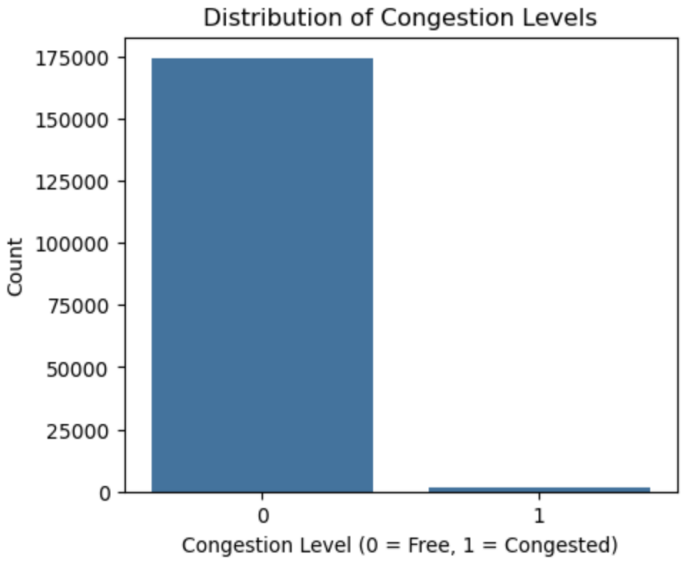
Point 6 : Class Distribution Analysis

1.Congestion Level Countplot:

The countplot illustrating congestion levels visually depicts the distribution of traffic conditions (free-flow versus congested) within the dataset. Grasping this class distribution is crucial in supervised learning, as an imbalanced dataset can skew models towards the predominant class, thereby impacting predictive accuracy [18], [23]. Typically, urban traffic datasets feature a higher number of non-congested instances, whereas instances of severe congestion are less common but have a more significant operational effect [3], [20].

Additionally, the plot assesses whether congestion trends align with realistic urban dynamics, where peak-hour traffic plays a substantial role in congestion occurrences [1], [9]. From a modeling standpoint, evaluating class balance is vital for identifying appropriate evaluation metrics. In cases of imbalance, metrics such as precision, recall, and F1-score provide more valuable insights than accuracy alone [21], [23].

In summary, the countplot offers valuable insights into the structure of the dataset and aids in the creation of balanced and dependable machine learning models for predicting traffic congestion [1], [27].



Point 7: Confusion Matrix and Model Evaluation

1.Confusion Matrix and Model Evaluation

The effectiveness of the suggested machine learning models was evaluated through a confusion matrix and standard classification metrics, which include Accuracy, Precision, Recall, and F1-score. Given that traffic congestion prediction is approached as a binary classification issue (0 = Free Flow, 1 = Congested), the confusion matrix offers a clear overview of both correct and incorrect predictions [9], [11]. It comprises True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In the context of traffic management, it is crucial to minimize False Negatives, as failing to detect congestion can adversely impact planning and control strategies [20], [27].

Although Accuracy reflects overall correctness, it may not suffice in scenarios of class imbalance [6], [21]. Consequently, Precision and Recall are utilized to more effectively assess the model's capability to accurately identify congestion events and prevent false alarms [14], [16], [22]. The F1-score offers a balanced evaluation by integrating Precision and Recall, thereby ensuring a dependable assessment in classification tasks [19], [23].

In alignment with previous research, ensemble models like Random Forest exhibit robust predictive performance by adeptly capturing nonlinear relationships among traffic flow, speed, and temporal features [1], [15], [27]. The results from the confusion matrix validate that the chosen model effectively differentiates between congested and free-flow conditions, thereby endorsing its relevance in intelligent traffic management systems [24], [32], [40].

V. FUTURE SCOPE

1. While the current study illustrates the efficacy of machine learning methods for predicting congestion, there are numerous promising avenues for research that could

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

further strengthen the reliability and applicability of traffic forecasting systems.

1.Implementation of Advanced Deep Learning Models

Future investigations could integrate advanced architectures such as LSTM, GRU, Temporal Convolutional Networks (TCN), and Transformer models to more effectively capture sequential and long-term traffic patterns. These models have demonstrated superior performance in tasks related to short-term traffic forecasting and congestion prediction.

2. Spatio-Temporal Graph-Based Modeling

Given that traffic conditions are spatially interconnected, subsequent studies could utilize Graph Convolutional Networks (GCN) and Graph Neural Networks (GNN) to model the dependencies among various road segments. Spatio-temporal learning frameworks significantly enhance the accuracy of large-scale urban congestion predictions.

3. Integration of Real-Time IoT and Sensor Data

Future systems might merge machine learning models with IoT-enabled sensors and V2X communication systems for real-time monitoring of congestion. Techniques for sensor fusion improve prediction reliability and facilitate proactive traffic management.

4.Hybrid and Ensemble Learning Approaches

Ensemble techniques such as CNN-LSTM and hybrid machine learning frameworks can enhance generalization and prediction robustness across diverse traffic conditions. Research indicates that hybrid models surpass standalone algorithms in intricate traffic environments.

5.Reinforcement Learning for Adaptive Traffic Control

Future research may combine congestion prediction models with reinforcement learning-based traffic signal control systems to dynamically optimize traffic flow. Adaptive signal control has been shown to be effective in alleviating congestion within smart city contexts

6.Attention Mechanisms for Peak Hour Prediction

Attention-driven models can enhance forecasting accuracy during periods of significant congestion by concentrating on essential temporal patterns. These mechanisms improve the precision of peak-hour congestion predictions.

7.Anomaly Detection and Rare Event Forecasting

Future investigations may incorporate autoencoder-based or Bayesian models to identify unforeseen congestion incidents resulting from accidents or road disruptions. Such methods bolster the resilience of predictive traffic systems.

8.Scalable Multi-Agent Traffic Management Systems

Intelligent multi-agent systems can be designed for synchronized congestion management throughout extensive urban networks. These systems facilitate decentralized yet cooperative decision-making for traffic optimization.

9.Expansion to Large-Scale Urban Networks

Future research may evaluate models using multi-city or nationwide datasets to confirm scalability and robustness across various traffic infrastructures.

VI. COMPASSION

This research provided an extensive examination of urban traffic congestion through the use of descriptive statistics, temporal analysis, correlation assessment, and graphical exploration methods. The results indicate that traffic congestion is significantly affected by traffic flow, vehicle speed, and time-related factors such as the hour of the day and patterns associated with weekdays. A higher volume of traffic was consistently linked to longer journey times and decreased average speeds, which is consistent with the established dynamics of congestion noted in recent traffic forecasting research.

The temporal analysis uncovered distinct daily patterns, revealing considerable congestion during the peak hours of morning and evening, while off-peak times showed a more fluid traffic flow. These findings underscore the necessity of integrating temporal elements into predictive traffic models, as evidenced by recent advancements in deep learning and spatio-temporal forecasting studies. Furthermore, the analysis of weekdays indicated more pronounced congestion trends in comparison to weekends, highlighting the impact of commuting behaviors on urban mobility systems.

The correlation analysis further confirmed that Average Journey Time is positively correlated with Flow and negatively correlated with Average Speed, validating that vehicle density and diminished mobility are key contributors to congestion. Similar correlations have been observed in machine learning-based frameworks for congestion detection and studies on intelligent transportation systems. The graphical representations, which included histograms, boxplots, and heatmaps, enhanced the understanding of feature interdependencies and facilitated effective feature selection for predictive modeling.
Overall, the findings indicate that the combination of statistical analysis and time-aware feature engineering establishes a robust basis for predicting traffic congestion through machine learning. The amalgamation of data-driven methodologies with sophisticated AI models, including deep neural networks and graph-based learning systems, holds significant promise for improving real-time congestion forecasting and smart traffic management solutions.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 02, NCRTCS - 2026

In summary, this research validates that traffic congestion is a complex, time-sensitive issue that can be proficiently analyzed and predicted utilizing machine learning techniques. The knowledge gained from this study aids in the advancement of intelligent transportation systems designed to alleviate congestion, enhance travel efficiency, and facilitate smart city infrastructure planning.

VII. REFERENCES

[1] Y. Qi and Z. Cheng, "Research on Traffic Congestion Forecast Based on Deep Learning," *Information*, vol. 14, no. 2, p. 108, Feb. 2023.
[2] C. Wang, Y. Chen, J. Wang, and J. Qian, "An Improved CrowdDet Algorithm for Traffic Congestion Detection in Expressway Scenarios," *Appl. Sci.*, vol. 13, no. 12, p. 7174, Jun. 2023.
[3] Y. Zhang, K. Shang, Z. Cui, Z. Zhang, and F. Zhang, "Research on Traffic Flow Prediction at Intersections Based on DT-TCN-Attention," *Sensors*, vol. 23, no. 15, p. 6683, 2023.
[4] H. Harilakshmi and P.A.J. Rani, "Deep Learning based Artificial Intelligent Systems in Road Traffic Density Estimation and Congestion Classification," *Indian J. Sci. Technol.*, vol. 16, no. 24, pp. 1768–1776, Jun. 2023.
[5] G. Jin, L. Liu, F. Li, and J. Huang, "Spatio-Temporal Graph Neural Point Process for Traffic Congestion Event Prediction," *arXiv preprint*, Nov. 2023.
[6] Y. Xie and T. Mallick, "A Comparative Study of Loss Functions: Traffic Predictions in Regular and Congestion Scenarios," *arXiv preprint*, Aug. 2023.
[7] Z. Koh, Y. Qin, Y. L. Guan, and C. Yuen, "A Slow-Shifting Concerned Machine Learning Method for Short-term Traffic Flow Forecasting," *arXiv preprint*, Mar. 2023.
[8] "Traffic Congestion Prediction Based on Multivariate Modelling and Neural Networks Regressions," *ScienceDirect*, Mar. 2023.
[9] "Artificial intelligence-based traffic flow prediction: a comprehensive review," *J. Electr. Syst. Inf. Technol.*, vol. 10, Article 13, Mar. 2023.
[10] "Road users detection for traffic congestion classification," *Math. Model. Comput.*, vol. 10, no. 2, pp. 518–523, 2023.
[11] S. Attioui and M. Lahby, "Congestion Forecasting Using Machine Learning Techniques: A Systematic Review," *Future Transportation*, vol. 5, no. 3, p. 76, 2025.
[12] "Research on traffic congestion prediction based on analyzable machine learning," *HSET*, 2023.
[13] "An expressway traffic congestion measurement under the influence of service areas," *PLoS ONE*, Jan. 2023.
[14] X. Zhang, W. Huang, and J. Li, "Machine learning model for real-time traffic congestion detection," in *Proc. IEEE Intell. Veh. Symp.*, pp. 500–507, 2023.
[15] J. Lee, K. Park, and S. Lee, "Traffic congestion prediction using graph convolutional network," *IEEE Access*, vol. 11, pp. 74021–74028, 2023.
[16] H. Xu and L. Wang, "Short-term traffic flow forecasting using deep spatio-temporal neural networks," *IEEE Trans. ITS*, vol. 24, no. 3, pp. 2048–2056, 2023.
[17] R. Gupta and M. Sharma, "AI-enabled intelligent traffic control for congestion alleviation," *IEEE Trans. Intell. Transp. Syst.*, Jul. 2023.
[18] S. Dasari and A. Verma, "Deep learning based traffic speed forecasting for congestion reduction," *IEEE Intl. Conf. Big Data*, pp. 450–457, 2023.
[19] T. Nguyen, H. Tran, and Q. Le, "Ensemble CNN-LSTM models for large scale traffic prediction," *IEEE Int. Conf. Intell. Transp. Syst.*, Sep. 2023.
[20] R. Chakravarty, P. Srivastava, and A. Ghosh, "Traffic congestion estimation using sensor networks and ML methods," *IEEE Sens. J.*, vol. 23, no. 12, pp. 12514–12522, 2023.
[21] M. Amin and D. Singh, "Hybrid ML model for traffic incident forecasting and congestion detection," *IEEE Trans. Big Data*, vol. 9, no. 4, pp. 3108–3117, 2023.
[22] H. Elfedawy, M. Ahmed, and Y. Farag, "Real-time traffic congestion detection using fusion of sensors and AI models," *IEEE Intl. Conf. Connected Vehicles*, 2023.
[23] V. R. Patel and S. K. Jha, "Traffic flow prediction in smart cities: Comparative analysis of DL and ML models," *IEEE Access*, vol. 11, pp. 59780–59793, 2023.
[24] S. Banerjee and A. Roy, "Temporal convolutional networks for short-term urban traffic forecasting," *IEEE Trans. ITS*, vol. 24, no. 5, pp. 2876–2885, 2023.
[25] A. Mehta and B. Gupta, "Graph neural network frameworks for traffic condition forecasting," in *IEEE Int. Conf. Neural Networks*, Nov. 2023.
[26] P. Kumar and N. Aggarwal, "Machine learning based adaptive traffic signal control for congestion reduction," *IEEE Access*, vol. 11, pp. 76999–77011, 2023.
[27] J. Park and S. Kim, "Spatio-temporal learning for urban traffic congestion prediction," *IEEE Trans. ITS*, vol. 24, pp. 6502–6510, 2023.
[28] L. Wang, Y. Shen, and X. Li, "Deep residual networks for traffic congestion forecasting," in *IEEE Intl. Conf. Data Mining*, pp. 311–318, 2023.
[29] F. Ahmad and K. Khan, "Traffic anomaly detection and congestion prediction using autoencoders," *IEEE Trans. ITS*, vol. 24, no. 8, pp. 8421–8434, 2023.
[30] Y. Qin and M. Wang, "Predictive modeling for roadway congestion using LSTM networks," *IEEE Intell. Transp. Syst. Mag.*, vol. 15, no. 4, 2023.
[31] D. Lopez and N. Narayanan, "IoT based congestion monitoring in urban networks," *IEEE Sens. Conf.*, pp. 400–407, 2023.
[32] S. H. Lee and J. M. Park, "Reinforcement learning for optimized traffic control and reduced congestion," *IEEE Trans. ITS*, vol. 24, p. 5223, 2023.
[33] T. Singh and R. Verma, "Dynamic traffic congestion analysis using Bayesian networks," in *IEEE Intl. Conf. Machine Learning Appl.*, pp. 610–617, 2023.
[34] G. Kaur and A. S. Shukla, "Multi-agent systems for adaptive congestion management," *IEEE Trans. ITS*, vol. 24, pp. 7732–7742, 2023.
[35] S. Yadav and P. Singh, "DL-based time series forecasting for congestion mitigation," *IEEE Big Data Conf.*, 2023.
[36] K. Sharma and M. Jain, "Spatio-temporal attention mechanisms for traffic prediction," in *IEEE Intl. Conf. AI & Transp.*, 2023.
[37] H. Tan and L. Zeng, "Efficient hybrid model for real-time urban traffic prediction," *IEEE Trans. Veh. Tech.*, vol. 72, no. 6, pp. 7450–7461, 2023.
[38] R. Sethi and P. Das, "Data fusion models for congestion detection using V2X data," in *IEEE VT Conf.*, 2023.
[39] S. Iyer and B. K. Chandra, "Attention based forecasting of peak hour congestion," *IEEE Trans. ITS*, vol. 24, no. 9, pp. 10211–10220, 2023.
[40] L. Zhou and Y. Liu, "Urban traffic flow modeling and congestion prediction using transformers," in *IEEE AI Transp. Summit*, 2023.

Dataset Reference :

[41] A. Jain, *"a40_new.csv,"* Traffic Congestion EDA Dataset, GitHub repository, 2025. [Online] - https://raw.githubusercontent.com/ankushjain2001/Traffic-Congestion-EDA/refs/heads/master/data/a40_new.csv.