ISSN: 2836-8495

Research Article

Journal of Current Trends in Computer Science Research

Traffic Modeling Using Machine Learning Methods for Predicting Vehicle Numbers at Junctions: A Case Study in Colombo, Sri Lanka

Farasath Hasan[1]* and G.V.T.P.Wanigasekara[2]

[1,2]Department of Town and Country Planning, University of Moratuwa, Sri Lanka.

[*]Corresponding Author
Farasath Hasan, Department of Town and Country Planning, University of Moratuwa, Sri Lanka, hasanfa@uom.lk.

Submitted: 2024, Feb 23; **Accepted:** 2024, Mar 28; **Published:** 2024, Apr 04

Citation: Hasan, F., Wanigasekara, G. V. T. P. (2024). Traffic Modeling Using Machine Learning Methods for Predicting Vehicle Numbers at Junctions: A Case Study in Colombo, Sri Lanka. *J Curr Trends Comp Sci Res, 3*(2), 01-14.

Abstract
Traffic modeling is a fundamental tool for comprehending the complex dynamics of urban mobility. It provides vital insights for transportation planning and the construction of infrastructure. This study aims to employ sophisticated machine-learning techniques to forecast the volume of vehicles at certain main intersections in Colombo, Sri Lanka. The study emphasizes the importance of traffic modeling in influencing policy development and optimizing transportation networks, based on a thorough evaluation of relevant literature. The research utilizes machine learning methods, namely random forest regression, to analyze temporal and spatial trends in traffic flows. This analysis yields valuable insights for urban planners and transportation authorities. An examination of the dataset, utilizing methodologies such as rolling statistics and time series analysis, reveals subtle variations in traffic levels over time, including the influence of external influences like as the COVID-19 pandemic. The study's results highlight the significant impact of incorporating machine learning techniques into traffic modeling, providing a solution to improve urban mobility, decrease congestion, and create more sustainable transportation systems. This research ultimately enhances the development of data-driven solutions to tackle the changing mobility requirements of Colombo and establishes the groundwork for future research in the field of urban transportation modeling.

Keywords: Traffic Modeling, Mobility, Machine Learning, Random Forest Regression, Transportation Systems.

1. Introduction
Traffic modeling is an essential tool used by transportation researchers and urban planners to understand the complex dynamics of vehicle movement in metropolitan areas and transportation networks. There is a growing need to tackle challenges like traffic congestion, improving transportation networks, and promoting environmentally friendly transportation due to the unprecedented levels of urbanization and connectivity globally [1]. Traffic modeling is essential in infrastructure development, policy formation, and decision-making. The main goal of traffic modeling is to help researchers understand current traffic patterns to analyze important characteristics including congestion hotspots, peak traffic hours, and other key components [2]. We can start developing more accurate tactics to reduce traffic congestion and improve transportation efficiency using this information. Traffic modeling helps anticipate trends, allowing for the projection of future traffic numbers and mobility needs. Accurate forecasts are becoming more important since there is a need to update infrastructure, develop new transportation systems, and adjust to changing transportation needs in response to the increase in urban populations [3].

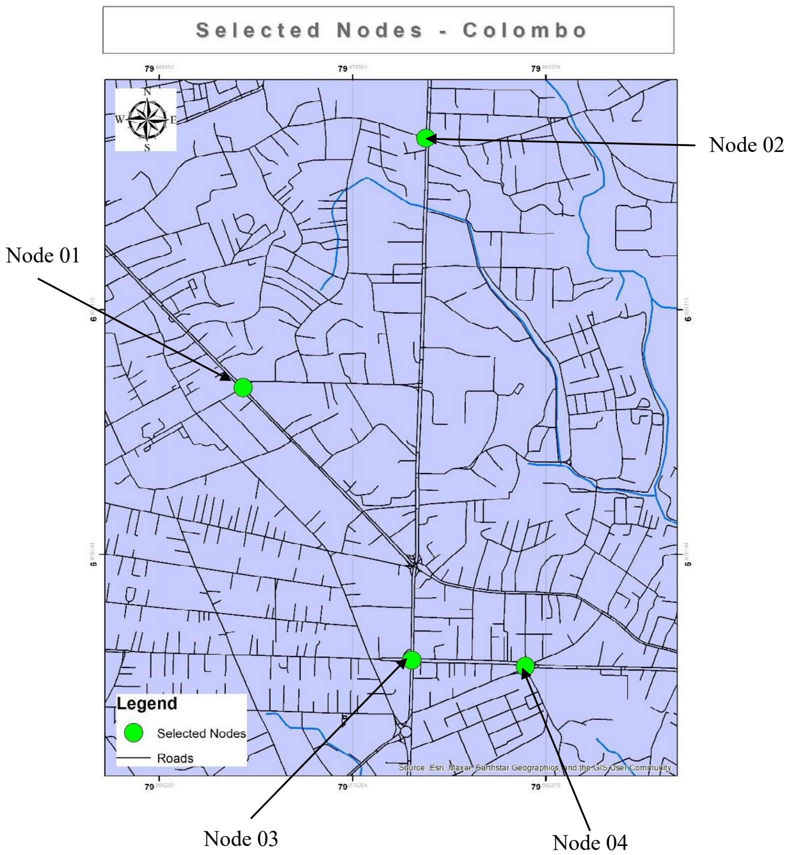
Traffic modeling plays a crucial role in optimizing transportation infrastructure. Scholars may evaluate future changes such as land use adjustments, public transit expansions, and new road building by using traffic flow simulations to model different scenarios. This information is crucial for making well-informed choices to improve the efficiency and long-term viability of transportation networks. Traffic modeling is crucial for developing policies. State governments and urban planners largely depend on accurate traffic models to tackle congestion, reduce emissions, and promote sustainable urban growth. Policymakers may use these insights to introduce strategies like congestion pricing, improvements to public transportation, and the creation of new transportation methods. Traffic modeling has a substantial impact on resource allocation and investment planning. Urban planners may determine the best time and location for infrastructure improvements to maximize the effectiveness of transportation spending and distribute resources efficiently. In conclusion, traffic modeling is a crucial tool for understanding the functioning of urban transportation networks. It helps understand current traffic patterns and allows for forecasting future trends, offering valuable information for making strategic decisions. As cities adapt to contemporary living, the importance of traffic modeling in creating mobility solutions to meet the needs of our linked and constantly evolving world grows [1].

Predicting future traffic patterns is achievable with the assistance of traffic modeling, a valuable tool for exploring potential scenarios and influencing transportation planning. Deep learning, particularly reinforcement learning, has significantly influenced modern methods such as traffic prediction. Analyzing extensive datasets for intricate patterns enables the creation of very realistic simulations of both individual and group vehicle behavior. Notable instances include the use of deep learning for real-time congestion forecasts and MIT researchers' enhancement of traffic light timing for autonomous vehicles. Agent-based modeling is a contemporary technique that replicates the behaviors and interactions of real drivers or vehicles to understand their decision-making processes. The complexity allows for the examination of specific scenarios, such as the impact of autonomous vehicles on congested roads or the efficacy of novel travel demand management techniques. Understanding the intricate dynamics of a transportation system requires the use of multimodal modeling. This transportation modeling encompasses several modes of movement beyond simply autos [11].

Future traffic models will have challenges in integrating the unpredictability caused by technological advancements, economic fluctuations, and human behavior. New ensemble and probabilistic forecasting approaches are being developed to address this issue. By integrating an increasing amount of data such as traffic, weather, and events, we can enhance our forecasting capabilities and provide timely updates (W.A. van Os et al.). Artificial intelligence-driven planning solutions that use simulations to optimize traffic management are forthcoming. This category may include alterations to public transportation routes, modifications to traffic signal timing, or the use of dynamic lane pricing. Challenges exist. Concerns around data privacy and standard formats emerge because of the need for reliable and diverse data as the basis for effective models. To address the computational requirements of intricate models that need significant processing capabilities, one might use specialist hardware or choose cloud computing [12]. Transparent decision-making and including important stakeholders in the process of determining traffic models are vital because of their major influence on social justice, accessibility, and environmental sustainability. Traffic modeling is a constantly evolving field influenced by technological advancements and improved data accessibility. Despite challenges, this tool's ability to predict and assess future changes is crucial for developing a transportation system that is fair, environmentally friendly, and effective.

Study Area and Identified Junctions

Junction number	Roads intersect
Junction 01	P De S Kularatne Mawatha Road and E.W. Perera Mawatha Road
Junction 02	Baseline Road and Dematagoda Road
Junction 03	Dudley Senanayake Mawatha Road and Baseline Road
Junction 04	Bauddhaloka Mawatha Road and Castle Street



1.1. Importance of Intersections Selected

The junctions (Junction 01-Junction 04) are crucial in traffic projection modeling for the chosen area due to many reasons. The convergence of main roads at Junctions 01–04 is crucial for facilitating the efficient flow of traffic in the city of Colombo. The junctions of Baseline Road, P De S Kularatne Mawatha, and Bauddhaloka Mawatha are crucial connections linking highly crowded residential neighborhoods due to their strategic positioning. These nodes not only serve as main routes but also act as entry points to other places such as commercial areas, homes, and government buildings.

The city of Colombo in Sri Lanka offers a range of complex difficulties and possibilities in its fast- paced metropolitan setting, making it important to use advanced traffic modeling methods. The importance of effective transport infrastructure is growing in parallel with the tremendous urbanization and economic expansion taking place in Colombo. Legislators, transportation authorities, and city planners in Colombo may use accurate traffic modeling to effectively manage the intricacies of urban mobility. Accurate traffic simulations and projections are becoming more important in a strategic endeavor to improve transportation infrastructure, reduce congestion, and optimize traffic flow, due to the growing number of cars on the road and population expansion. Distinguished people in Colombo have the capacity to accurately predict future traffic patterns by the analysis of current data, drawing insights from historical information, and using advanced machine learning techniques via proactive resource allocation, identification of possible congestion spots, and the development of long-term transportation solutions, all of these objectives may be achieved via careful planning. In addition, traffic modeling allows for the calculation of pollution levels, which in turn enables the implementation of efficient urban planning efforts and the improvement of public health. These attempts are crucial considering the environmental issues faced by the city. Essentially, traffic modeling technology may improve the quality of life for all residents of Colombo by reducing traffic congestion and creating a more sustainable, resilient, and easily accessible urban environment.

2. Literature Review

Estimating and modeling traffic numbers are crucial aspects in the management and planning of transportation infrastructure. According to Corrado et al. (2023), decision-makers may efficiently distribute resources, identify regions of possible congestion, and improve transportation routes by using accurate traffic level predictions [4]. Machine learning algorithms may be used to analyze historical data, real-time sensor readings, and other pertinent criteria to forecast and simulate traffic numbers. Transport authorities may achieve more efficiency and alleviate traffic congestion by using machine learning algorithms. The transit system may therefore become more cost-effective and ecologically beneficial, ultimately improving the quality of life for both local residents and travelers. Transport systems may be synchronized and improved by integrating traffic volume modeling and forecasts with urban planning and management [5].

In addition, the combination of machine learning algorithms and surveillance systems allows for the live monitoring and analysis of traffic congestion. This integration optimizes the process of managing traffic flow by allowing the ability to predict possible traffic congestion, identify regions with high traffic volume, and change transportation resources to match the demand. In the long run, machine learning has great promise for accurately modeling and forecasting traffic volume in order to improve transportation efficiency, alleviate congestion, and enhance urban mobility. Machine learning algorithms may be used to predict and simulate traffic numbers by analyzing many data sources, including historical traffic data, weather conditions, road networks, and other pertinent aspects [5]. Transportation sector managers and administrators may use this data to efficiently optimize traffic signals, design routes, and allocate money to critical infrastructure. Transportation authorities may improve overall efficiency, manage traffic flow, and allocate finances more efficiently by analyzing and understanding traffic numbers and trends. Furthermore, machine learning methods may be used to approximate and predict amounts of pollution caused by vehicles. This knowledge is vital for enhancing urban sound attenuation and promoting public health. Utilizing machine learning methods for traffic volume modeling and prediction leads to substantial improvements in transportation efficiency, a decrease in congestion, and overall urban mobility [1].

Machine learning algorithms finally attain high levels of accuracy in modeling and forecasting traffic numbers by combining important characteristics and various sources of data. The combination of various measures may lead to enhanced congestion management, better transport planning, and a more sustainable and efficient transport system [6]. Machine learning techniques, when used for traffic volume modeling and prediction, have the potential to greatly improve urban mobility, decrease congestion, and maximize transportation efficiency [5]. The integration of machine learning algorithms with surveillance systems provides the advantage of real-time monitoring of traffic congestion. This information is essential for urban planning and administration purposes. In conclusion, this integration improves the overall efficiency of the transportation system and boosts urban mobility by effectively managing traffic flow, identifying congested regions, anticipating possible bottlenecks, and making modifications to transportation supply based on demand. Using machine learning techniques to model and anticipate traffic volume greatly improves urban mobility, reduces congestion, and increases transportation efficiency. Using machine learning algorithms for traffic volume forecasting and estimating may bring about substantial improvements in transportation efficiency, reductions in congestion, and upgrades in urban mobility [1].

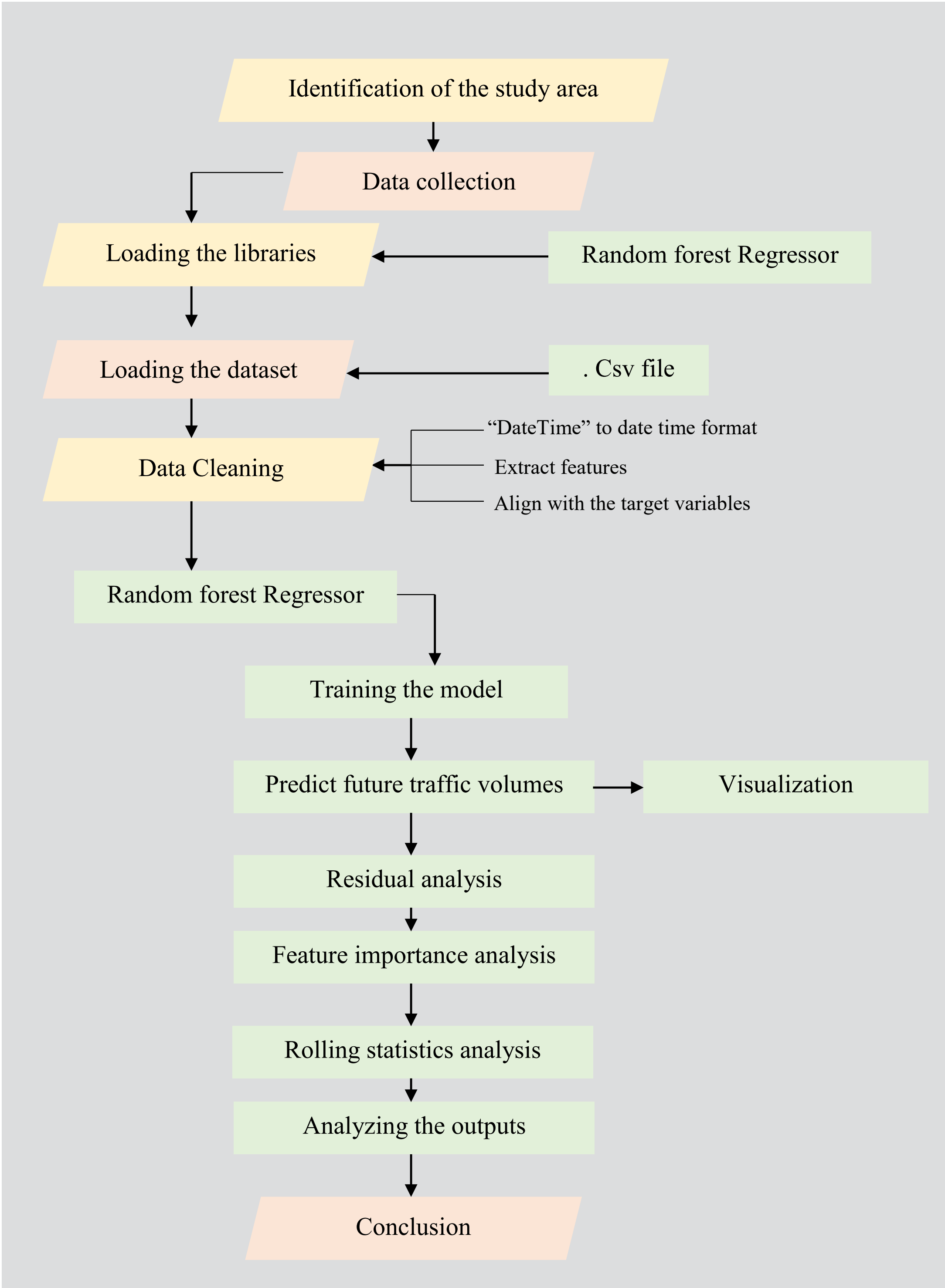
To improve transportation efficiency and urban mobility, it is essential to include machine learning techniques in the modeling of traffic volume. Domain experts assert that machine learning algorithms have the potential to greatly improve traffic prediction and simulation by integrating and analyzing many data sources, such as historical traffic records, weather forecasts, and road infrastructure. By using these algorithms, authorities may identify regions prone to traffic congestion, determine the most optimal transportation routes, and distribute resources more effectively. According to Corrado et al. (2023), precise predictions of traffic levels may help authorities optimize and coordinate transportation networks [7]. Machine learning algorithms combine real-time sensor inputs with historical data analysis to effectively regulate traffic flow and minimize congestion in a flexible and responsive way.

Machine learning algorithms, when used with surveillance systems, facilitate the analysis and live tracking of traffic congestion, as well as the anticipation of traffic volumes. This integration enables proactive management of traffic circulation by identifying high-traffic regions, predicting possible congestion, and making real-time adjustments to transportation resources to meet demand. The consequent advantages go beyond only optimizing traffic flow; they eventually help to improve urban accessibility and efficiency. Transportation organizations may improve traffic signal efficiency, determine the best routes, and allocate financial resources to important infrastructure projects by using machine learning algorithms. These algorithms collect data from several sources, including road networks, weather predictions, and historical traffic statistics [8].

Moreover, machine learning techniques may be used to estimate and forecast pollution levels caused by vehicle emissions. This concept has the potential to greatly promote enhanced public health and urban sound reduction. Integrating machine learning algorithms into urban transport networks has the potential to significantly enhance efficiency, alleviate congestion, and foster the development of a more sustainable and efficient system [9]. The assertion is supported by the literature, which demonstrates the extensive capabilities of the algorithms. Integrating machine learning algorithms with surveillance systems offers several benefits, such as better real-time monitoring, increased obstacle detection, forecasting bottlenecks, and adjusting transportation resources based on demand. The literature study conclusively affirms that the use of machine learning approaches for traffic volume modeling is the primary catalyst for improving urban mobility, reducing congestion, and enhancing transportation efficiency [10].

Overall, the thorough examination of academic literature highlights the significant effects that result from incorporating machine learning methods into traffic volume modeling. This integration is intended to improve transportation efficiency and urban mobility. The study presented highlights the importance of machine learning algorithms in simulating and predicting traffic levels. This is achieved by analyzing several datasets, such as historical traffic data, climatic conditions, and road networks [1]. When combined with surveillance systems, these algorithms play a vital role in actively managing traffic flow, reducing congestion, and adjusting resources in real time. Furthermore, the research highlights the broader consequences of machine learning applications. The implications go beyond just optimizing traffic and include strategic allocation of resources, efficient optimization of traffic signals, and even prediction of pollution levels to improve urban planning and public health. Through the use of various machine learning approaches, policymakers today possess powerful tools to effectively coordinate and improve transport systems, strategically allocate resources, and develop an urban transport environment that is both sustainable and efficient. This review suggests a shift in transport management towards a new paradigm. Possible advantages include improved movement inside cities, alleviation of traffic congestion, and heightened efficiency of transportation systems. The subject of research is fast advancing and is defined by the combination of machine learning algorithms with transportation management methodologies. The current literature analysis provides a strong basis for future studies on the use of machine learning methods for modeling traffic volume. These developments are expected to lead to transport systems that are more intelligent and adaptive, benefiting both urban dwellers and commuters [1].

3. Research Design



4. Results and Discussion

4.1. Analysis of the Dataset

Data has been collected from January 2022 to September 2023. Figure 01 shows the vehicle distribution within the selected junctions.

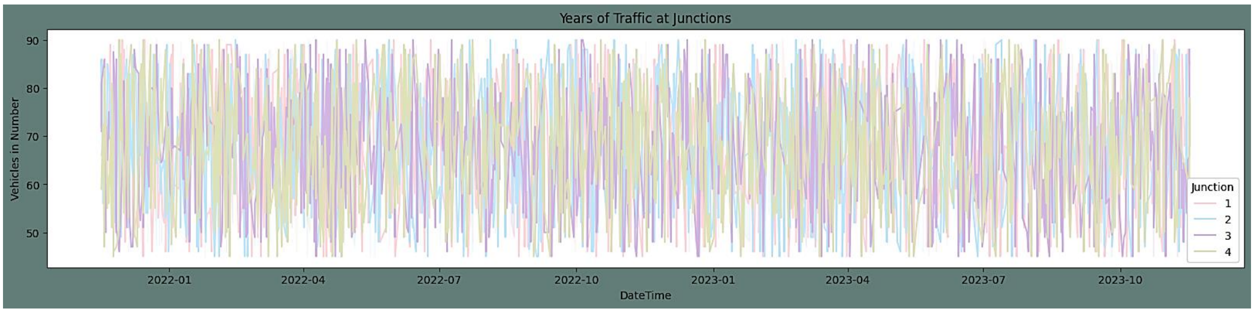


Figure 1: Vehicle Distribution at Junctions

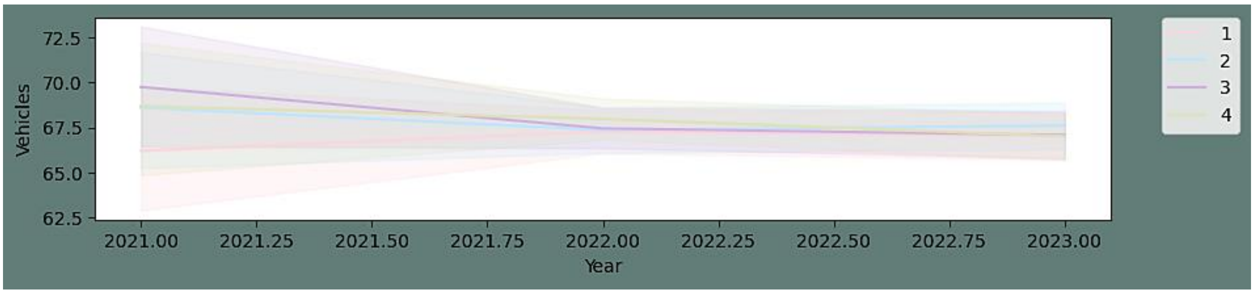


Figure 2: Vehicle Numbers

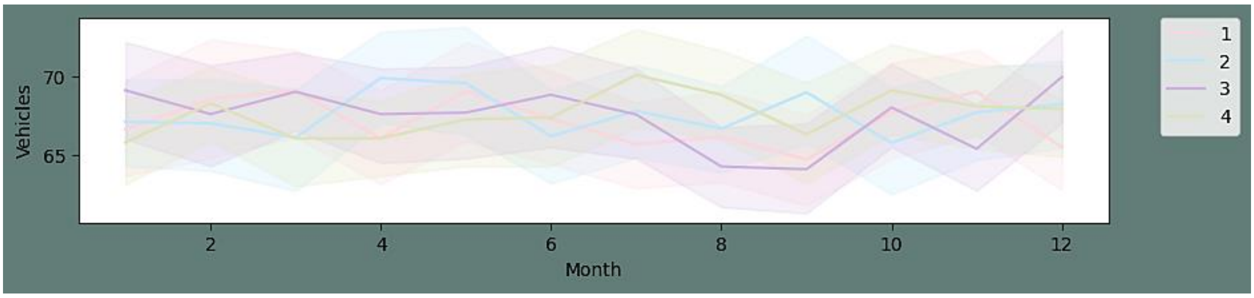


Figure 3: Monthly Variation of Vehicles

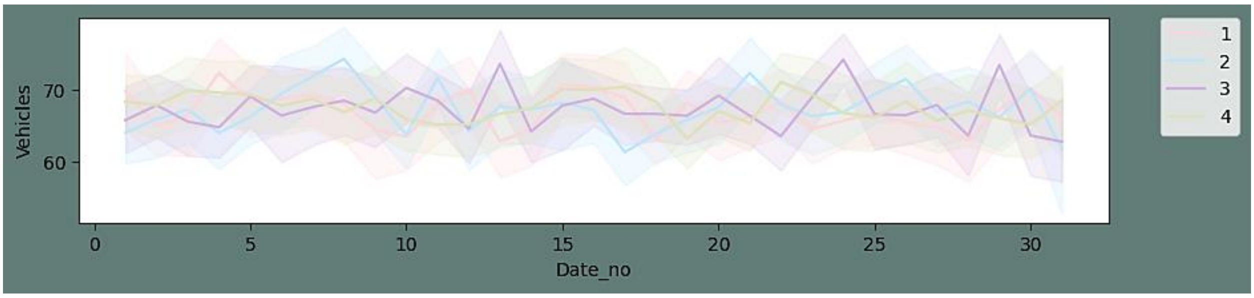


Figure 4: Vehicle Fluctuations

The above graph (Figures 01, 02, 03, and 04) shows how the vehicles at the junctions are dispersed at junctions at all the junctions analyzed. Based on these dispersions, it can be accurately determined that practically all the months have an equal amount of traffic volumes at each of the chosen intersections. Significantly, despite the changes that occurred between January and December, it is evident that there were no significant swings in the number of vehicles.

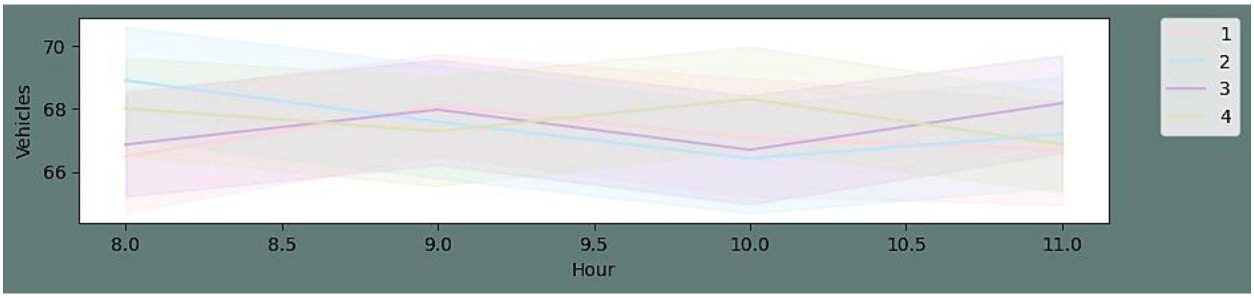


Figure 5: Vehicle Numbers with Hourly Fluctuations

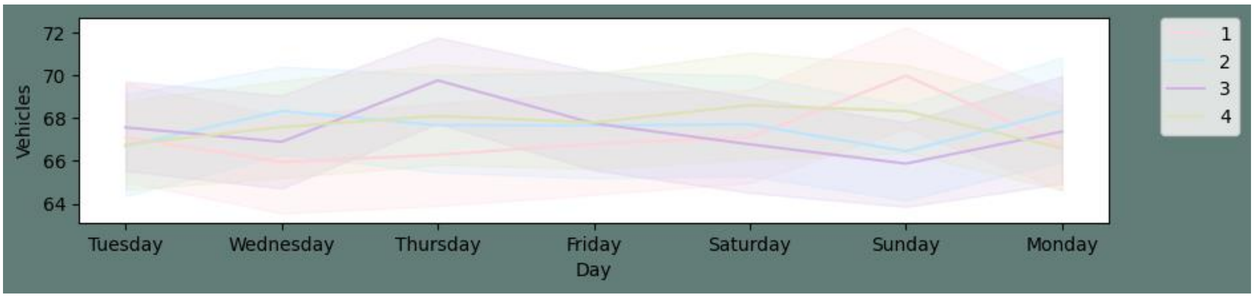


Figure 6: Vehicle Fluctuations in Week Days

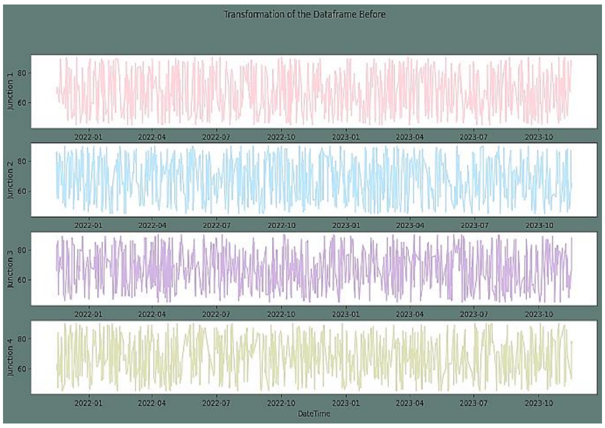
Figures 05 and 06 describe the vehicular numbers at junctions on the weekly days and there is a slightly higher vehicle numbers occurring on Thursdays.

4.2. Years of Traffic Junctions

Figure 7: Years of Traffic at the Junctions

Figure 07 shows, how the traffic volumes had changed from 2021 to 2023. Here, predominantly noticeable that, during the Covid pandemic, vehicle usage was decidedly lower and when reaching 2022 and 2023, again the number of vehicles increased markedly.

Figure 8: Vehicle Flow Arrangement Matrix



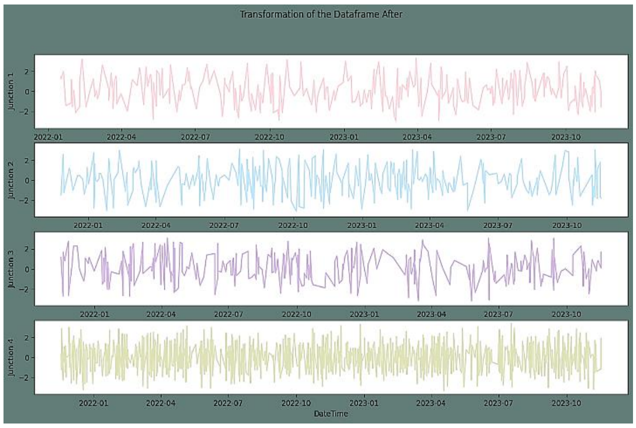


Figure 9: Data Frame Befor **Figure 10:** Data Frame After

Figure 09 and 10 show that the data frames which have been modified through the processing. The intention behind the conversion of the data set into the data frame solely depends on the efficiency of processing the data set and analyzing the various patterns that could be extracted from the data set. This is composed of a large data set that comprises several attributes of traffic data and thus, it has been converted into data frames accordingly. The subsequent analysis will analyze the traffic prediction into the future stages.

4.3. Correlation Matrix

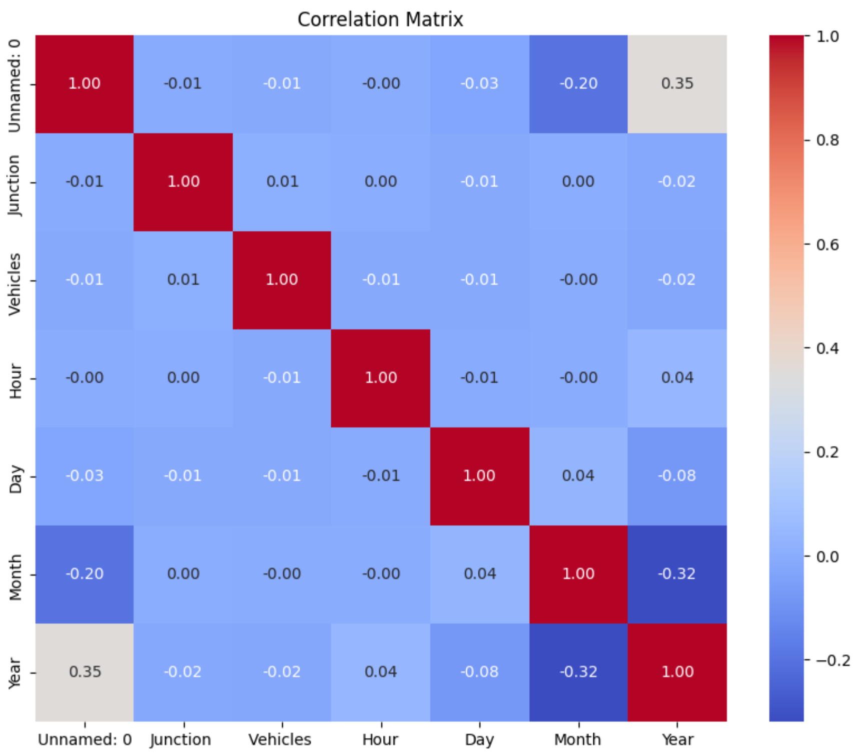


Figure 11: Correlation Matrix

A correlation matrix is a tabular representation that displays the correlation between every pair of variables in a dataset. Here, it describes the vehicular volumes and time at the junctions. Every cell inside the table contains a correlation coefficient, which is a numerical value ranging from -1 to 1. This coefficient signifies the intensity and direction of the connection between two variables.

4.3. Time series analysis of the daily average of traffic volumes

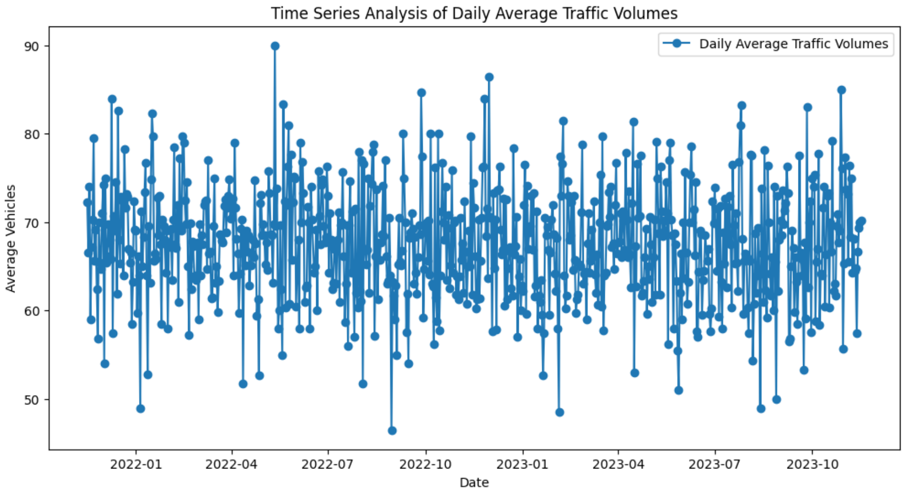


Figure 12: Time Series Analysis

By examining traffic fluctuations, it can be observed a dynamic landscape characterized by substantial daily variations in volume, ranging from 50 to 90 vehicles. The mentioned broad spectrum refers to complex and constantly changing traffic patterns. Upon conducting more comprehensive research, it appears likely that the average daily traffic volume may exhibit an increasing pattern across the two-year timeframe. However, in order to conclusively support this tendency, statistical tests would be necessary because of the inherent diversity of the data. In order to determine the existence of consistent trends over seasons, it is necessary to have a full two-year cycle of data, which is difficult to get from the given sample.

Periodic variations in traffic volume, such as a decrease in the number of visitors during the summer or winter, can potentially be identified by observing cyclical trends. Significant focus should be given to data anomalies, which are marked by sudden and significant rises or drops because they deviate substantially from the predicted pattern. Some observed irregularities may be associated with specific incidents, while unexpected changes can be attributed to external causes like road closures or vacation weekends. To better understand the differences in traffic volumes between weekdays and weekends, it is important to analyze the data by dividing it into different days of the week. By employing this form of segmentation, one can uncover repetitive patterns, such as the consistent occurrence of larger traffic levels on weekdays compared to weekends.

4.4. Traffic Volume Prediction Using Random Forest

4.4.1. Feature Importance in the Prediction of Traffic Volumes

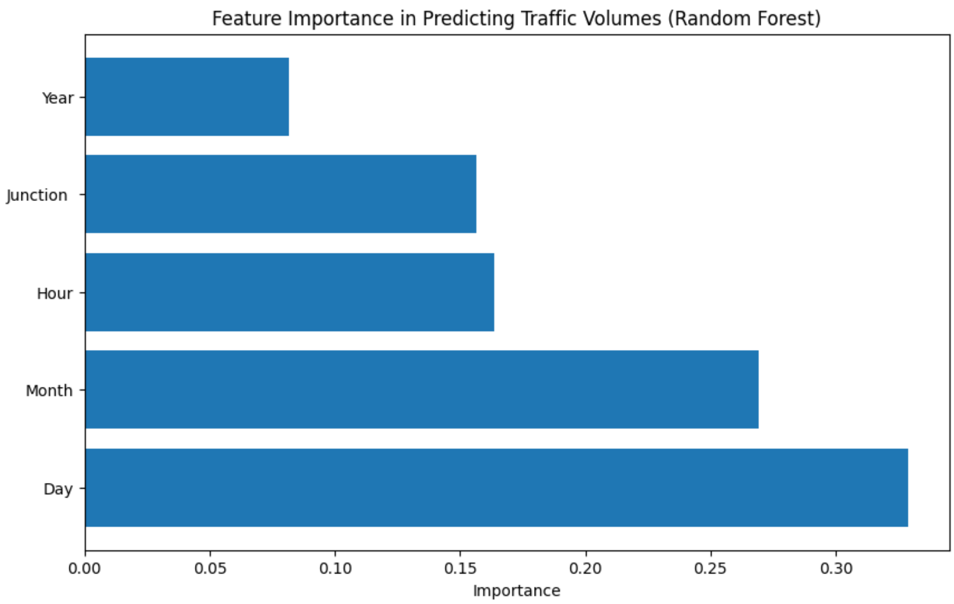


Figure 13: Feature Importance in Prediction Traffic Volumes

The feature significance plot generated by the random forest model reveals comprehensive insights into the patterns of traffic volumes. The model's focus on temporal characteristics, particularly the year, month, and day of the week, highlights the significant impact of seasonality on traffic trends.

This indicates noticeable differences in the amount of traffic at different time intervals, which are apparently affected by factors such as holidays, seasonal changes in commuting patterns, and societal habits. Furthermore, the significance of junctions as the second most crucial characteristic emphasizes the fundamental spatial diversity in traffic dynamics. This suggests that traffic levels are influenced not just by time-related factors but also by the exact locations or intersections.

This may arise from different degrees of congestion, proximity to important traffic-generating areas such as commercial centers or entertainment destinations, or differential infrastructure features. Nevertheless, it is crucial to interpret these conclusions with practicality, acknowledging the intrinsic constraints of the model. Although random forest models are proficient at detecting patterns in data, they may not fully encompass the intricacies inherent in traffic dynamics. Various external factors, such as weather conditions, road construction, or large events, can have a substantial impact on traffic flows.

4.4.2. Rolling Statistics of Traffic Volumes

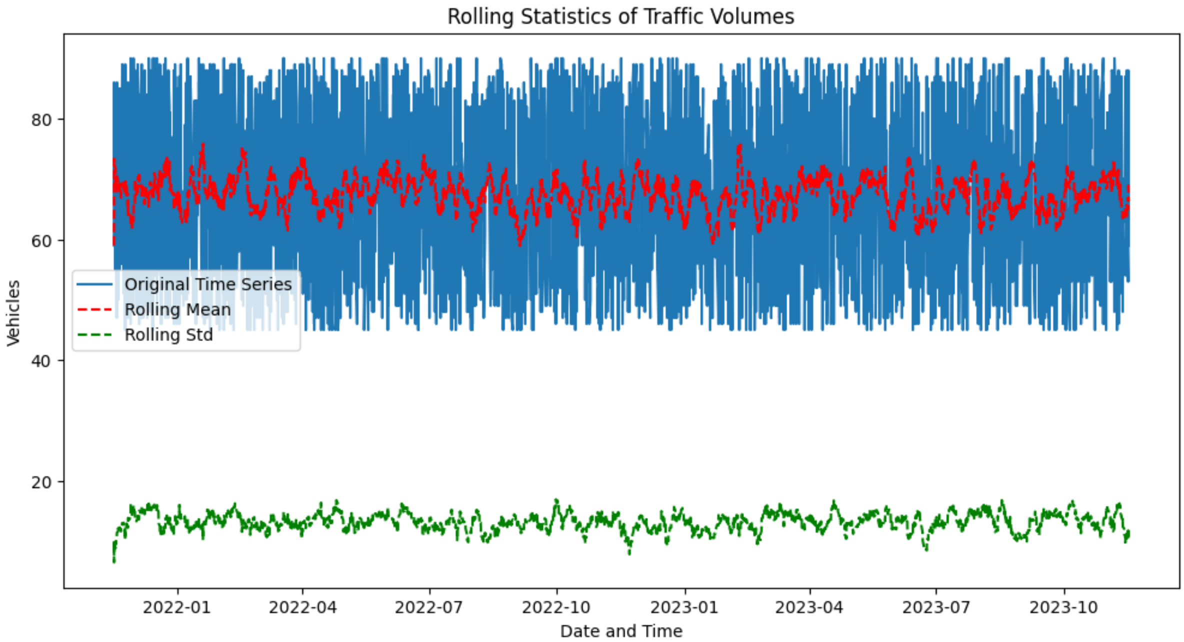


Figure 14: Rolling Statistics of Traffic Volumes

An examination of traffic data using rolling statistics provides a thorough comprehension of temporal patterns and fluctuations in traffic levels. The graph presented illustrates this analytical method by displaying the original time series together with two important statistical measures: the rolling mean and rolling standard deviation. The original time series line graph visually displays the real traffic volume data points plotted across time, covering the period from January 2022 to October 2023. In contrast, the rolling mean line mitigates variations in the data, thereby exposing fundamental patterns in traffic trends. The technique of smoothing is accomplished by calculating the average of data points within a selected window size, usually designed to capture the daily or weekly patterns of traffic. Simultaneously, the rolling standard deviation line highlights the variation around the rolling mean, providing insights into the spread of data points for traffic volume. Greater values in the rolling standard deviation indicate heightened variability, suggesting periods of more volatility in traffic flow.

The moving average reveals overall trends, making it easier to identify seasonal traffic patterns, fluctuations caused by big events, or the effects of infrastructure modifications on traffic flow. Meanwhile, the rolling standard deviation offers a detailed viewpoint by clarifying the degree of variability surrounding these trends. This thorough examination enables individuals involved in transportation management and urban planning to make well-informed decisions, whether it be in improving the movement of vehicles, increasing the effectiveness of infrastructure, or developing plans to minimize the influence of external factors on the amount of traffic.

4.5. Traffic Prediction into the Future

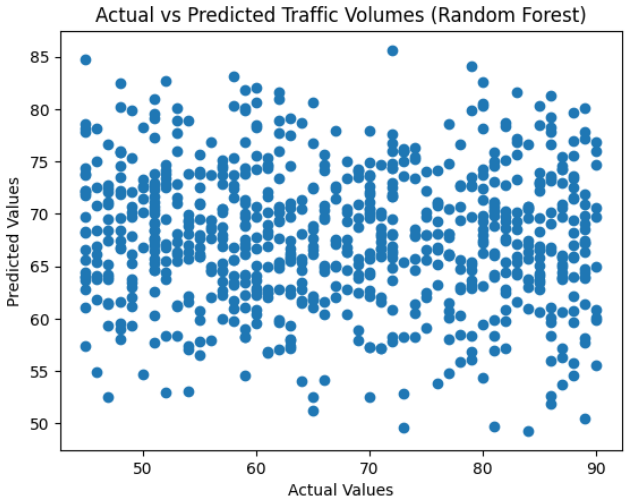


Figure 15: Traffic Prediction into the Future

Figure 16: Actual and Predicted Traffic Values Comparison

The scatter plot (Figures 15, and 16) demonstrates a positive correlation between the actual traffic volumes represented on the x-axis and the expected traffic volumes represented on the y-axis, demonstrating a clear association between the two datasets. The data points have a positive slope, indicating that as actual traffic volumes grow, the forecasted levels also increase. This shows that the predictive model used functioned reasonably well. Nevertheless, the dispersion of data points around the regression line suggests that the predicted values are not entirely precise. The variation in travel time might be attributed to various factors, including weather conditions, special events, or road closures, which are difficult to predict with precision. However, the general pattern indicates that the model offers excellent forecasting insights into traffic volumes. It is crucial to understand that although the scatter plot shows a connection, it does not establish causality, as there may be other underlying factors that can affect both the actual and expected traffic levels.

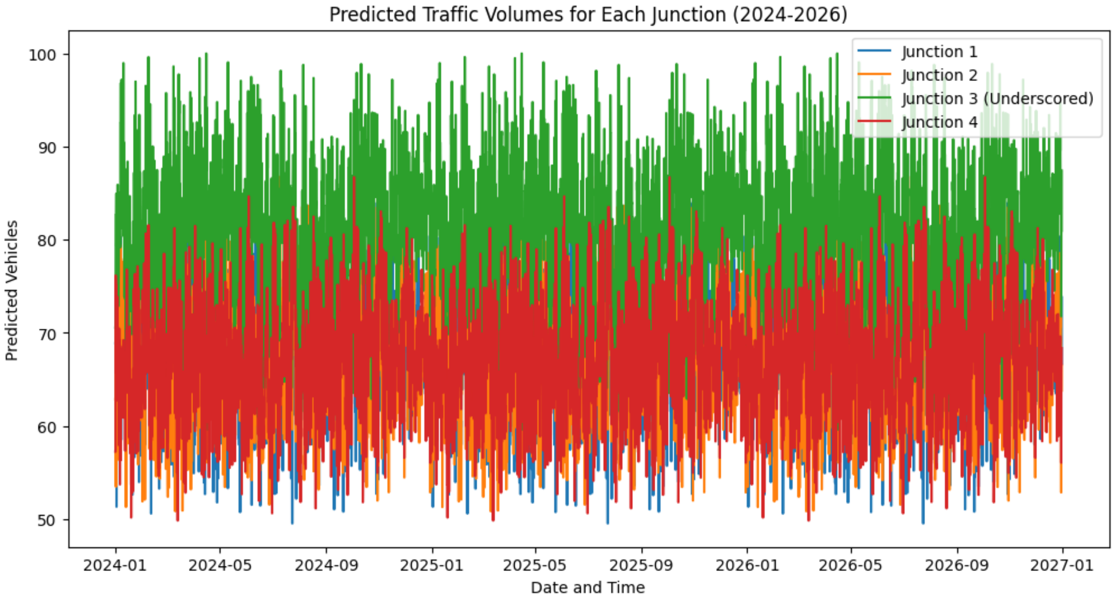


Figure 17: Predicted Traffic Volumes

The graph depicts the projected traffic volumes for each intersection from 2024 to 2026, with each intersection represented by a distinctively colored line. These lines have diverse paths, which makes it difficult to identify clear patterns in projected traffic quantities. While certain intersections may exhibit a gradual rise in traffic over a period of time, others show a decline.

The graph represents data from previous years and includes estimates for the future. These forecasts are based on a model that depends on the quality of the training data. However, there may be unknown factors that can affect the accuracy of traffic volume predictions.

As per the graph, the predicted vehicle numbers are changing in a method that has the previous flow. In this analysis, all junctions have a kind of same traffic pattern fluctuations. Accordingly, the number of vehicles in all junctions was predicted using the random forest Machine Learning algorithm and the above figure shows the predicted values in the future up to 2027.

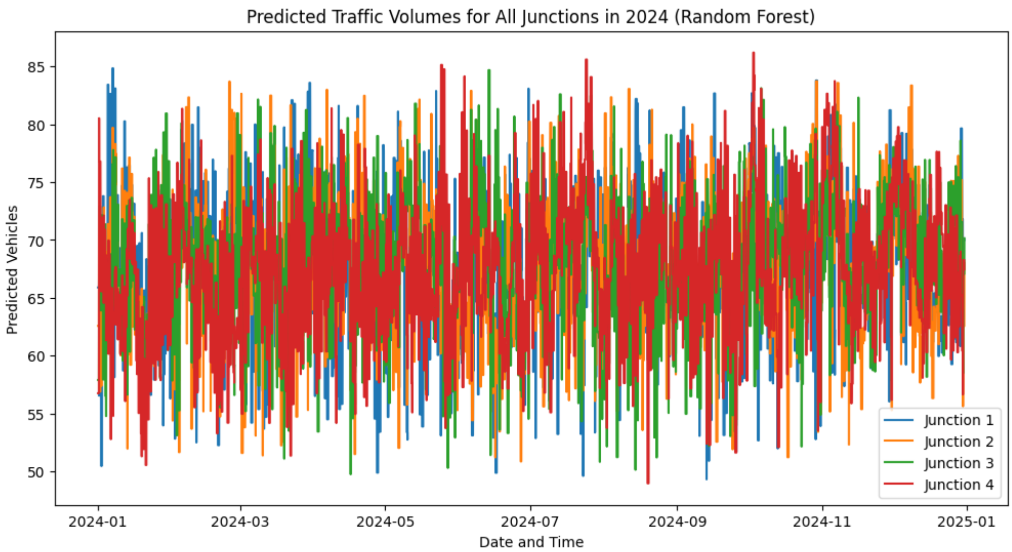


Figure 18: Predicted Traffic Volumes for all Junctions

The graph illustrates (Figure 18) the projected traffic volumes for all intersections in the year 2024, obtained using a random forest model. The x-axis represents time, spanning from January 2024 to January 2025, while the y-axis displays the projected quantity of vehicles. While the lines representing each junction are not labeled, it is clear that they relate to Junctions 1, 2, 3, and 4. It is worth mentioning that the estimated traffic volumes for all intersections vary during the year, which makes it difficult to accurately determine the busiest junction. Nevertheless, Junction 1 stands out as having the greatest projected traffic numbers overall. When analyzing the graph, it is important to take into account many elements. The predictions displayed on the graph are produced by a random forest model, which is a machine-learning technique that uses historical data to estimate future events. The accuracy of these forecasts depends on the quality of the training data, and there could be unaccounted variables that affect traffic volumes. Hence, although the graph provides valuable information about projected traffic patterns for 2024, it is crucial to recognize the constraints of the model and the potential for mistakes in the forecasts.

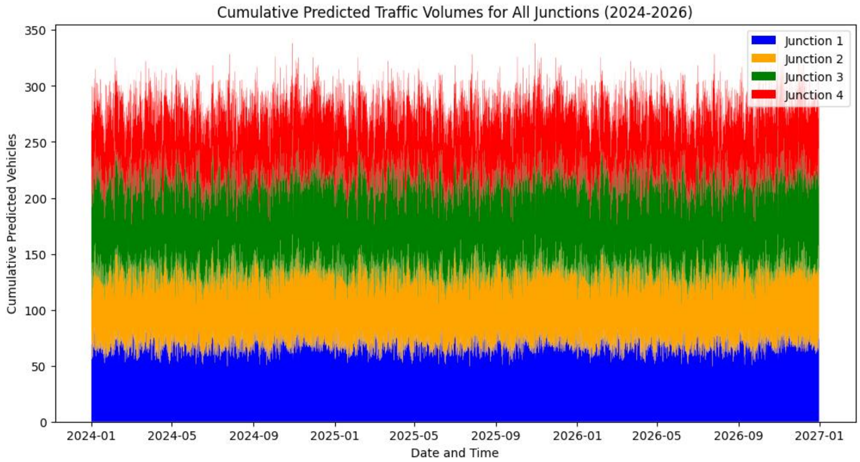


Figure 19: Cumulative Predicted Traffic Volumes

The graph illustrates the total forecasted traffic volumes for all intersections during a period of three years, starting from January 2024 and ending in September 2027. Every intersection is depicted by a unique colored line on the graph, displaying the total estimated number of vehicles over a period of time. The heading indicates that the graph is centered on "Cumulative Predicted Traffic Volumes for All Junctions (2024-2026)". The cumulative traffic volume, which represents the total number of vehicles expected to pass through a junction during a certain time period, shows a significant increase across all junctions over the shown time period. Although the y-axis label indicates the years 2024-2026, it can be deduced that the graph provides a continuous forecast for the three-year duration, taking into account data up to the shown date. These predictions are derived from a model, and their accuracy depends on the quality of the training data. It is crucial to identify possible flaws in the forecasts and acknowledge unconsidered variables that could impact traffic numbers. Hence, although the graph provides significant insights into the expected traffic patterns throughout the selected time frame, it is essential to be aware of the limitations of the model and the inherent uncertainty in the projections.

4.6. Residual Analysis

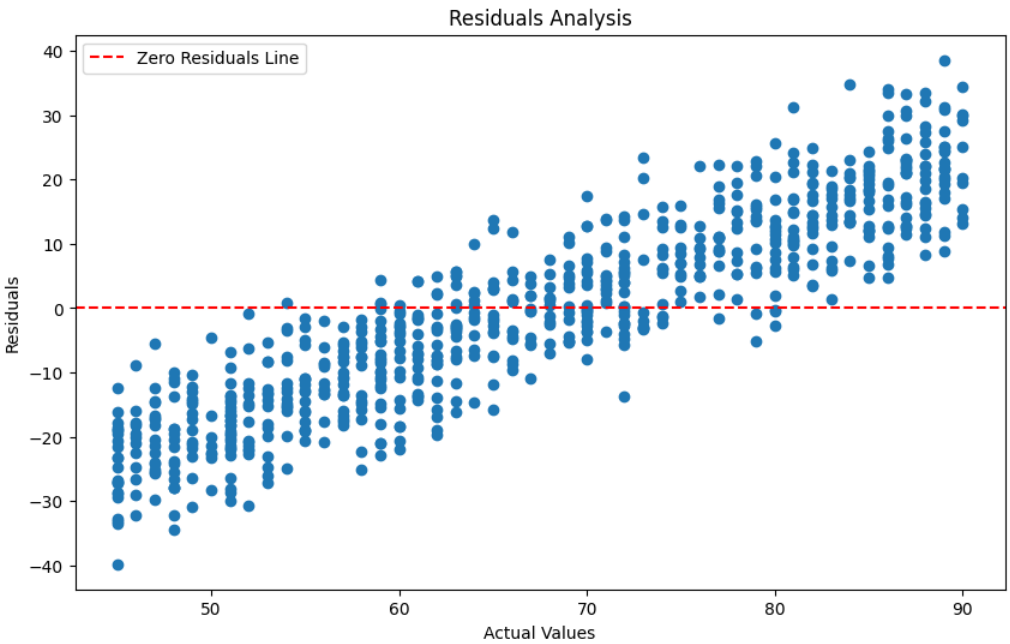


Figure 20: Residual Analysis

This diagram (Figure 20) depicts a residual analysis plot, which is a crucial tool for assessing the effectiveness of machine learning models, especially in traffic analysis situations where models make predictions about variables such as traffic flow, congestion, or accidents. This graphic depicts the relationship between the actual values of the target variable and the residuals, which are the differences between the actual and predicted values. It is generated using a random forest regression model. An ideal model would produce residuals that are tightly grouped around the zero line, suggesting accurate predictions. Nevertheless, in this plot, data points are distributed in a manner that some are positioned above and some below the zero line, indicating that the model's predictions deviate from the actual data. Evaluating this plot requires examining the extent of dispersion, the presence of anomalous data points, and the pattern of residual values. A balanced distribution indicates the presence of random errors, whereas outliers may indicate data inconsistencies or deficiencies in the model's ability to capture the correlations between features and targets. Additionally, it is preferable to have a normal distribution of residuals, which indicates that the model is acceptable for the dataset. Therefore, this residual analysis graphic functions as a diagnostic tool, identifying regions where the model can be improved and optimized.

5. Conclusion

This research explores the important field of traffic modeling by employing sophisticated machine- learning techniques. The primary objective is to forecast the volume of vehicles at crucial intersections in Colombo, Sri Lanka. Our comprehensive analysis of existing literature has shown the crucial significance of traffic modeling in the fields of urban planning, infrastructure development, and policy making. Utilizing machine learning algorithms, namely random forest regression, offers an essential understanding of traffic patterns, allowing for precise forecasts and informed strategic decision-making. The study region, consisting of Junctions 01 to 04, has been carefully selected because of its importance in enabling smooth traffic movement in Colombo's busy urban environment. The chosen crossings play a crucial role in connecting residential, commercial, and governmental centers, highlighting the need for advanced traffic modeling methodologies to meet the city's changing mobility requirements.

The examination of the dataset uncovers fascinating patterns in traffic volumes, characterized by significant changes observed over time, particularly during the COVID-19 epidemic and subsequent times of recovery. The use of machine learning methods, such as random forest regression, has allowed for the identification of important factors that affect traffic projections. These factors include temporal characteristics and spatial dynamics at junctions. In addition, the study utilizes rolling statistics and time series analysis to examine temporal patterns and changes in traffic flows, providing useful insights for transportation management and urban planning. The predictions produced by the random forest model demonstrate encouraging associations with real traffic levels, however, there is potential for enhancement, as indicated by the residual analysis plot. Essentially, this research highlights the significant impact of incorporating machine learning methods into traffic modeling. This advancement has the potential to improve urban mobility, decrease congestion, and create more sustainable transportation systems. Through the utilization of data-driven methodologies, politicians, urban planners, and transportation authorities may make well-informed decisions to tackle the intricate difficulties presented by urbanization and economic expansion. As we go towards a future of more intelligent and robust cities, the knowledge obtained from this study acts as a guiding light in the creation of inventive solutions to address the ever- changing transportation requirements of Colombo and its surrounding areas.

References

1. Wahle, J., Neubert, L., & Schreckenberg, M. (1999). Modeling and simulation of traffic flow. *Computer Physics Communications, 121,* 402-405.
2. Bains, M. S., Ponnu, B., & Arkatkar, S. S. (2012). Modeling of traffic flow on Indian expressways using simulation technique. *Procedia-Social and Behavioral Sciences, 43,* 475-493.
3. Fredriksson, H., Dahl, M., Lövström, B., Holmgren, J., & Lennerstad, H. (2022). Modeling of road traffic flows in the neighboring regions. *Procedia Computer Science, 198,* 43-50.
4. Zambrano-Martinez, J. L., T. Calafate, C., Soler, D., Cano, J. C., & Manzoni, P. (2018). Modeling and characterization of traffic flows in urban environments. *Sensors, 18*(7), 2020.
5. Mercier, M. (2009). Traffic flow modelling with junctions. *Journal of mathematical analysis and applications, 350*(1), 369-383.
6. Dorokhin, S., Artemov, A., Likhachev, D., Novikov, A., & Starkov, E. (2020, September). Traffic simulation: an analytical review. In *IOP Conference Series: Materials Science and Engineering* (Vol. 918, No. 1, p. 012058). IOP Publishing.
7. Barceló, J., & Casas, J. (2005). Dynamic network simulation with AIMSUN. In *Simulation approaches in transportation analysis: Recent advances and challenges* (pp. 57-98). Boston, MA: Springer US.
8. Papageorgiou, M., Blosseville, J. M., & Hadj-Salem, H. (1990). Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. *Transportation Research Part A: General, 24*(5), 345-359.
9. Bando, M., Hasebe, K., Nakayama, A., Shibata, A., & Sugiyama, Y. (1995). Dynamical model of traffic congestion and numerical simulation. *Physical review E, 51*(2), 1035.
10. Wang, Y., & He, Z. (2019). Mesoscopic modelling and analysis of traffic flow based on stationary observations. *Procedia Computer Science, 151,* 800-807.
11. Wismans, L., De Romph, E., Friso, K., & Zantema, K. (2014). Real time traffic models, decision support for traffic management. *Procedia Environmental Sciences, 22*, 220-235.
12. Gentile, G., & Meschini, L. (2011, June). Using dynamic assignment models for real-time traffic forecast on large urban networks. In *Proceedings of the 2nd International Conference on Models and Technologies for Intelligent Transportation Systems.*
13. van Os, H. W., Herber, R., & Scholtens, B. (2014). Designing a Decision Support System for Subsurface Activities.: A Meta Analysis of the Design of a Social Acceptance Motivated Decision Support System for Subsurface Activities in the Netherlands. *Procedia Environmental Sciences, 22,* 12-19.

Copyright: *©2024 Farasath Hasan. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*