

TRAFFIC PREDICTION USING MACHINE LEARNING AND REAL TIME DATA

Anil Jaiswal, UG Student, Computer Science and Engineering, Integral University

Ashahad Khan, UG Student, Computer Science and Engineering, Integral University

Mohammad Ashraf, UG Student, Computer Science and Engineering, Integral University

Supervisor and corresponding author: Mrs shra fatima,

Abstract

Traffic congestion has emerged as a significant challenge in modern urban environments, primarily driven by rapid population growth, increased vehicle ownership, and inadequate infrastructure planning. Conventional traffic management systems often rely on static or historical data, which limits their ability to adapt to dynamic and real-time traffic conditions, resulting in inefficiencies, delays, and increased environmental pollution. This research presents an advanced intelligent traffic prediction framework that integrates machine learning techniques with real-time data streams to enhance prediction accuracy and decision-making capabilities. The proposed system leverages historical traffic datasets in conjunction with real-time inputs such as weather conditions, road occupancy, vehicle density, and temporal variations to capture complex traffic patterns. A combination of machine learning models, including traditional regression methods and advanced deep learning approaches such as Long Short-Term Memory (LSTM) networks, is employed to forecast traffic flow with high precision.

The study also incorporates data preprocessing, feature engineering, and model optimization techniques to improve system performance and robustness. Experimental results indicate that the inclusion of real-time data significantly reduces prediction errors

and enhances model responsiveness to sudden traffic fluctuations caused by incidents, weather changes, or peak-hour variations.

The proposed framework demonstrates strong potential for deployment in Intelligent Transportation Systems (ITS) and smart city infrastructures, enabling efficient traffic management, reduced congestion, improved travel time estimation, and enhanced urban mobility. Furthermore, the system can support decision-makers, urban planners, and commuters by providing reliable, real-time traffic insights and route optimization strategies.

Keywords: Traffic Prediction, Machine Learning, Real-Time Data, Intelligent Transportation Systems (ITS), Deep Learning, LSTM, Traffic Flow Forecasting, Smart Cities, Data Analytics, Urban Mobility, Time-Series Analysis.

1. Introduction

With the rapid evolution of smart city initiatives and digital infrastructure, traffic management has emerged as a critical research domain within Artificial Intelligence (AI) and data science. The exponential growth in urban populations, coupled with the increasing number of vehicles, has intensified traffic congestion in metropolitan areas worldwide. This congestion not only leads to prolonged travel time and commuter frustration but also contributes significantly to environmental pollution, fuel wastage, and economic losses. As cities continue to expand, the need for intelligent, adaptive, and data-driven traffic management systems has become more essential than ever before.

Traditional traffic prediction models primarily depend on static historical datasets and conventional statistical approaches, which are often inadequate for capturing the dynamic

and non-linear nature of real-world traffic conditions. These models lack the capability to respond effectively to sudden changes such as accidents, weather variations, road blockages, and peak-hour fluctuations. As a result, there is a growing demand for advanced predictive systems that can process real-time information and provide accurate, timely insights into traffic flow patterns.

Machine Learning (ML), a core subset of Artificial Intelligence, offers a powerful solution to these challenges by enabling systems to learn from large volumes of both historical and real-time data. ML algorithms can identify complex patterns, correlations, and temporal dependencies in traffic datasets, thereby improving prediction accuracy. The integration of real-time data sources, including Internet of Things (IoT) sensors, Global Positioning System (GPS) devices, mobile applications, and traffic surveillance cameras, further enhances the adaptability and responsiveness of traffic prediction models. These data streams provide continuous updates on vehicle density, road occupancy, speed variations, and environmental conditions, allowing for dynamic and context-aware forecasting. The proposed approach focuses on improving traffic flow estimation, reducing congestion, and supporting intelligent decision-making for urban transportation systems. By leveraging advanced analytical methods and data-driven insights, this study contributes toward the development of scalable and sustainable solutions for smart city environments, ultimately enhancing urban mobility, road safety, and overall quality of life.

2. Literature Review

Recent advancements in traffic prediction research emphasize the increasing role of Machine Learning (ML) and data-driven methodologies in addressing the complexities of

urban traffic systems. Early traffic forecasting approaches primarily relied on statistical and mathematical models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA), which were effective for linear time-series analysis. However, these traditional models exhibited significant limitations in handling the non-linear, dynamic, and stochastic nature of real-world traffic data, particularly under rapidly changing environmental and road conditions. With the emergence of machine learning techniques, researchers began exploring more adaptive and scalable models capable of capturing complex traffic patterns. Algorithms such as Support Vector Machines (SVM), Random Forests, k-Nearest Neighbors (k-NN), and Artificial Neural Networks (ANN) have demonstrated improved prediction accuracy by learning from large-scale historical datasets. These models are capable of identifying hidden relationships among variables such as traffic volume, time intervals, road conditions, and external influencing factors. In particular, ensemble methods like Random Forests have shown robustness against overfitting and improved generalization performance in traffic flow prediction tasks.

The evolution of deep learning has further transformed the landscape of traffic prediction systems. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have gained widespread attention due to their ability to effectively model temporal dependencies in sequential data. LSTM networks overcome the limitations of traditional neural networks by maintaining long-term memory, making them highly suitable for time-series forecasting problems such as traffic flow prediction. Furthermore, hybrid deep learning architectures combining Convolutional Neural Networks (CNN) and LSTM have been proposed to simultaneously capture spatial and

temporal features of traffic data. CNN layers are used to extract spatial correlations from traffic networks, while LSTM layers handle temporal dynamics, resulting in more accurate and robust predictions.

Recent studies have also explored the use of big data analytics and cloud computing frameworks to process and manage large volumes of traffic data efficiently. The combination of real-time data processing with advanced machine learning and deep learning models has significantly enhanced the performance of traffic prediction systems. Despite these advancements, challenges such as data heterogeneity, scalability, and model interpretability remain areas of ongoing research. Overall, the literature indicates that the integration of machine learning with real-time data analytics is a promising direction for developing intelligent, adaptive, and efficient traffic management solutions in smart city environments.

3. Methodology

The proposed traffic prediction framework adopts a hybrid data-driven approach that integrates both historical datasets and real-time data streams to achieve accurate and dynamic traffic forecasting. This approach is designed to overcome the limitations of traditional static models by incorporating continuously evolving traffic conditions into the prediction process. The dataset used in this study consists of multiple attributes, including traffic volume, timestamps, road conditions, vehicle density, and environmental factors such as weather conditions. Historical data is obtained from publicly available repositories such as Kaggle, while real-time data is acquired through Application Programming Interfaces (APIs) and IoT-enabled sources, ensuring continuous updates and real-world relevance.

The initial phase of the methodology involves comprehensive data preprocessing, which plays a crucial role in enhancing model performance. Raw data collected from multiple sources is often noisy, incomplete, and inconsistent; therefore, it undergoes several preprocessing steps, including data cleaning, normalization, and transformation into a structured format suitable for machine learning algorithms. Missing values are addressed using interpolation and imputation techniques to preserve temporal continuity, while outliers are detected and removed using statistical methods to maintain data integrity. Feature engineering is also performed to extract meaningful attributes such as peak hours, day-of-week patterns, and seasonal variations, which significantly contribute to improving prediction accuracy.

Following preprocessing, the dataset is partitioned into training and testing subsets to facilitate model development and evaluation. Multiple machine learning models are implemented to analyze and compare prediction performance. Linear Regression is employed as a baseline model due to its simplicity and interpretability, providing a reference for evaluating more advanced techniques. Ensemble learning methods such as Random Forest are utilized to capture non-linear relationships and improve generalization by aggregating multiple decision trees. Furthermore, deep learning models, particularly Long Short-Term Memory (LSTM) networks, are incorporated to effectively model temporal dependencies in sequential traffic data. LSTM networks are well-suited for time-series forecasting as they retain long-term memory and capture complex temporal patterns that traditional models often fail to represent.

A key component of the proposed system is the integration of real-time data streams into the prediction pipeline. Real-time inputs, such as live traffic updates, weather conditions,

and road incidents, are continuously fed into the trained model through APIs, enabling dynamic and adaptive prediction. This real-time integration allows the system to respond promptly to sudden fluctuations in traffic conditions, thereby enhancing prediction reliability and practical applicability in real-world scenarios.

4. System Architecture

The proposed system architecture is designed as a scalable and modular framework that integrates data acquisition, processing, model development, and real-time prediction to support intelligent traffic management. The architecture consists of four interconnected components: data collection, data processing, model training, and prediction, each contributing to the overall efficiency and adaptability of the system in dynamic urban environments.

The first component, data collection, involves gathering traffic-related information from diverse sources to ensure both historical depth and real-time responsiveness. Historical datasets are obtained from publicly available repositories, while real-time data is acquired through IoT-enabled sensors, GPS devices, traffic surveillance cameras, and weather monitoring systems. This multi-source data acquisition enables the system to capture spatial and temporal variations in traffic conditions, including vehicle density, speed, road occupancy, and environmental factors.

The second component, data processing, is responsible for transforming raw and heterogeneous data into a structured and usable format. This stage includes data cleaning, normalization, feature extraction, and integration of multiple data streams. Noise removal, handling of missing values, and synchronization of time-series data are essential tasks performed in this module. Additionally, feature engineering techniques are applied to

derive meaningful attributes such as peak-hour indicators, traffic trends, and contextual variables, which enhance the learning capability of the prediction models.

The third component, model training, focuses on building and optimizing machine learning and deep learning models using preprocessed data. Various algorithms, including Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks, are trained to learn underlying traffic patterns. The training process involves parameter tuning, validation, and performance evaluation to ensure robustness and generalization. The inclusion of deep learning models enables the system to capture complex temporal dependencies and non-linear relationships inherent in traffic data.

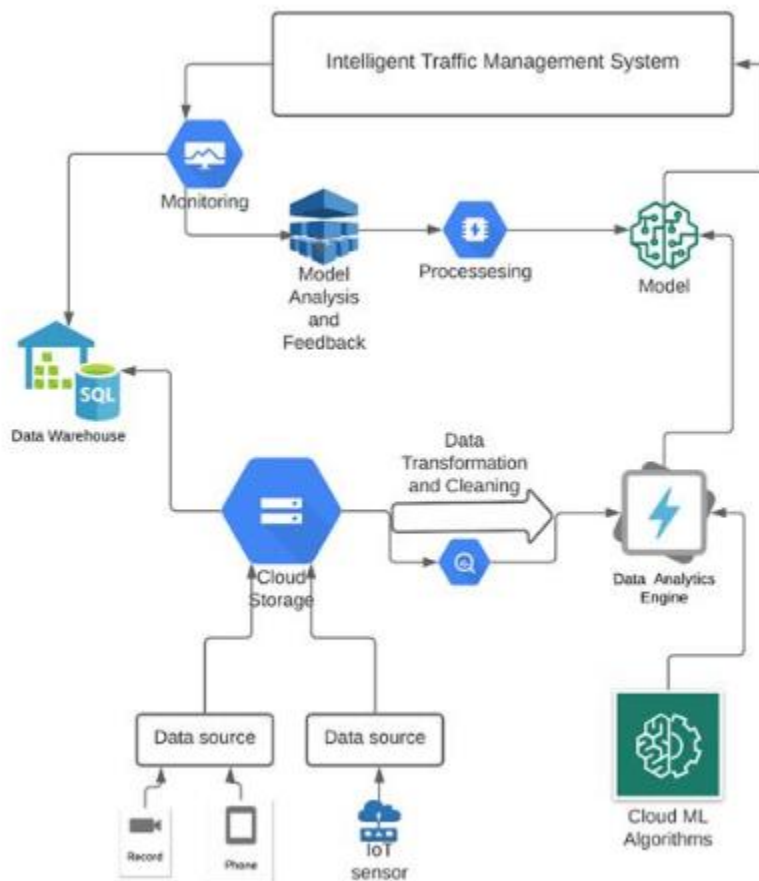


Figure 1: Proposed System Architecture for Real-Time Traffic Prediction Using ML

The final component, prediction and deployment, integrates the trained model with real-time data streams to generate continuous traffic predictions. The system dynamically updates its outputs based on incoming live data, allowing it to respond effectively to sudden changes such as congestion, accidents, or weather disruptions. The predicted results are presented through a user-friendly interface that provides real-time traffic status, congestion alerts, estimated travel time, and optimal route recommendations.

This architecture ensures high scalability, as it can accommodate large volumes of data and support distributed processing using cloud or edge computing technologies. Furthermore, its modular design allows easy integration with existing Intelligent Transportation Systems (ITS) and smart city infrastructures. By enabling real-time adaptability and accurate forecasting, the proposed system architecture significantly contributes to efficient traffic management, reduced congestion, and improved urban mobility.

5. Results and Discussion

The experimental evaluation of the proposed traffic prediction framework demonstrates that machine learning models integrated with real-time data significantly outperform traditional approaches that rely solely on historical datasets. The inclusion of dynamic inputs such as live traffic updates, weather conditions, and road occupancy enables the model to adapt to rapidly changing traffic scenarios, thereby improving overall prediction accuracy and robustness. Among the implemented models, the Long Short-Term Memory (LSTM) network exhibits superior performance due to its ability to effectively capture temporal dependencies and sequential patterns inherent in traffic flow data. Unlike

conventional regression and ensemble methods, LSTM maintains long-term memory, allowing it to learn complex variations across time intervals. This capability results in lower prediction errors, as reflected in evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

The study also highlights the effectiveness of hybrid approaches that combine multiple data sources and modeling techniques. By leveraging both spatial and temporal features, the system provides more reliable and context-aware predictions. The analysis reveals that the proposed framework successfully identifies critical traffic patterns, including peak traffic hours, high-density zones, and congestion-prone routes. Additionally, the system offers optimized route suggestions, which can assist commuters in minimizing travel time and avoiding congested areas.

Conclusion

This research presents a robust and scalable traffic prediction framework that effectively combines machine learning techniques with real-time data integration to address the challenges of urban traffic congestion. By utilizing both historical datasets and continuous real-time inputs, the proposed system achieves improved prediction accuracy, adaptability, and reliability compared to conventional methods. The incorporation of advanced deep learning models, particularly LSTM networks, enables the system to capture complex temporal dynamics and deliver precise traffic forecasts.

The proposed framework demonstrates strong potential for real-world deployment in Intelligent Transportation Systems (ITS) and smart city infrastructures. Its ability to provide real-time traffic insights, congestion alerts, and optimal route recommendations can significantly contribute to reducing travel time, minimizing fuel consumption, and

enhancing overall road safety. Moreover, the system supports informed decision-making for urban planners, transportation authorities, and commuters, thereby promoting efficient and sustainable urban mobility. Despite its promising results, certain challenges remain, including data heterogeneity, scalability issues, and the need for high computational resources for real-time processing. Future research can address these limitations by incorporating more advanced techniques such as Graph Neural Networks (GNN) for modeling spatial dependencies in traffic networks and reinforcement learning for adaptive traffic signal control. Additionally, the integration of edge computing and Internet of Things (IoT) infrastructure can further enhance real-time processing capabilities and reduce system latency.

References

1. Yisheng, L., et al. "Traffic Flow Prediction with Big Data: A Deep Learning Approach." *IEEE Transactions on Intelligent Transportation Systems*.
2. Williams, B. L., & Hoel, A. "Modeling Vehicular Traffic Flow Using ARIMA." *Journal of Transportation Engineering*.
3. Ma, X., et al. "Learning Traffic as Images: A Deep CNN Approach." *IEEE Transactions on Intelligent Transportation Systems*.
4. Hochreiter, S., & Schmidhuber, J. "Long Short-Term Memory." *Neural Computation*, 1997.
5. Zheng, Z., et al. "Urban Traffic Prediction Using Deep Learning." *Elsevier Journal*.
6. Lv, Y., et al. "Traffic Flow Prediction with Big Data: A Deep Learning Approach." *IEEE TITS*, 2015.
7. Polson, N. G., & Sokolov, V. O. "Deep Learning for Short-Term Traffic Flow Prediction." *Transportation Research Part C*, 2017.
8. Zhang, J., Zheng, Y., & Qi, D. "Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction." *AAAI*, 2017.
9. Yu, B., Yin, H., & Zhu, Z. "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework." *IJCAI*, 2018.
10. Li, Y., Yu, R., Shahabi, C., & Liu, Y. "Diffusion Convolutional Recurrent Neural Network." *ICLR*, 2018.

11. Ke, J., et al. "Short-Term Forecasting of Passenger Demand Under On-Demand Ride Services." *IEEE TITS*, 2017.
12. Huang, W., Song, G., Hong, H., & Xie, K. "Deep Architecture for Traffic Flow Prediction." *IEEE TITS*, 2014.
13. Wu, Y., Tan, H. "Short-Term Traffic Flow Forecasting with Spatial-Temporal Correlation." *PLOS ONE*, 2016.
14. Min, W., & Wynter, L. "Real-Time Road Traffic Prediction with Spatio-Temporal Correlations." *Transportation Research Part C*, 2011.
15. Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. "Short-Term Traffic Forecasting: Overview." *Transportation Research Part C*, 2014.
16. Chen, C., et al. "LSTM Neural Network for Traffic Flow Prediction." *Procedia Computer Science*, 2016.
17. Fu, R., Zhang, Z., & Li, L. "Using LSTM for Traffic Flow Prediction." *IEEE ITSC*, 2016.
18. Zhang, Z., et al. "A Deep Learning Approach for Traffic Data Prediction." *IEEE ITS Conference*, 2017.
19. Wang, H., et al. "Traffic Speed Prediction Using Deep Learning Method." *IEEE Access*, 2019.
20. Guo, S., et al. "Attention Based Spatial-Temporal Graph Convolutional Networks." *AAAI*, 2019.
21. Shi, X., et al. "Convolutional LSTM Network for Precipitation Nowcasting." *NIPS*, 2015.
22. Yu, H., et al. "Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction." *Neurocomputing*, 2017.
23. Sun, S., et al. "Traffic Flow Forecasting Using Hybrid Deep Learning Methods." *IEEE Access*, 2020.
24. Abbas, S. H., et al. "Machine Learning-Based Smart Traffic Prediction Models." *Springer*, 2023.
25. Akhtar, W., et al. "AI-Based Traffic Prediction for Smart Cities." *Elsevier*, 2022.
26. Koesdwiady, A., et al. "Improving Traffic Flow Prediction Using Hybrid Methods." *IEEE TITS*, 2016.