International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

# An Intelligent Machine Learning Framework for Public Transport Passenger Demand Forecasting Using Time Series and Regression Models

**M.Swathi[1], Srishti Kulkarni[2]**

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

**Abstract.** Accurate prediction of passenger demand is essential for improving the efficiency, reliability, and resource management of urban public transportation systems. Reliable forecasting enables transport authorities to optimize vehicle scheduling, reduce passenger waiting times, and enhance the overall quality of service. This study presents a comparative machine learning framework for forecasting passenger demand across stations of Lima Metro Line 1 using five predictive algorithms: Prophet, Linear Regression, Random Forest, Decision Tree, and Auto-Regressive Integrated Moving Average (ARIMA). Historical passenger records collected from the official OSITRAN transportation database are utilized to train and evaluate the forecasting models. Prior to model development, the dataset undergoes comprehensive preprocessing, including data integration, cleaning, datetime transformation, and passenger count aggregation to ensure consistency and reliability. The trained models are assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination ($R^2$) to determine their predictive capability. Experimental analysis demonstrates that the Prophet model consistently achieves the highest forecasting accuracy, producing the lowest prediction errors and the highest $R^2$ value among all evaluated models. Although Decision Tree and Linear Regression require considerably less computational time, they exhibit lower predictive performance compared with Prophet. The findings indicate that Prophet effectively captures temporal demand patterns and seasonal variations present in passenger transportation data, making it a practical and reliable solution for intelligent public transport planning, demand forecasting, and operational decision support in modern urban transit systems.

**keywords:** Passenger Demand Prediction, Public Transportation, Machine Learning, Prophet Model, ARIMA, Random Forest, Decision Tree, Linear Regression, Time Series Forecasting, Smart Transportation.

## I. Introduction

Rapid urbanization and continuous population growth have significantly increased the demand for efficient public transportation systems across metropolitan cities worldwide. As urban populations expand, transportation authorities face growing challenges in managing passenger flow, minimizing congestion, improving service reliability, and utilizing available resources efficiently. Public transportation plays a vital role in reducing traffic congestion, lowering environmental pollution, and supporting sustainable urban development. Consequently, accurately forecasting passenger demand has become an important research topic that enables transportation agencies to improve scheduling, optimize fleet allocation, and provide better services to commuters [1], [2].

Passenger demand within urban transportation systems is influenced by numerous dynamic factors, including travel time, station location, working hours, weekends, holidays, seasonal variations, and unexpected events. These factors produce complex temporal patterns that are difficult to capture using traditional statistical forecasting techniques. Inaccurate demand estimation often leads to overcrowded vehicles during peak periods or underutilized transportation resources during off-peak hours, both of which negatively affect operational efficiency and passenger satisfaction. Therefore, intelligent forecasting systems capable of analyzing historical travel patterns are increasingly being adopted to support effective transportation planning [3].

Traditional forecasting approaches, such as statistical regression models and classical time-series methods, have been widely employed to estimate passenger demand. Although these techniques provide acceptable results for relatively simple datasets, they generally struggle to model highly nonlinear relationships and changing temporal behaviors observed in modern transportation systems. Furthermore, many conventional models require strong assumptions regarding data distribution and seasonality, limiting their applicability in real-world environments where passenger behavior continuously evolves. These limitations have motivated researchers to investigate machine learning algorithms that automatically learn complex patterns from historical transportation data without relying on rigid statistical assumptions [4].

Recent advances in artificial intelligence and machine learning have significantly transformed transportation analytics by enabling accurate forecasting of passenger flow using historical travel records. Algorithms such as Linear Regression, Decision Tree, Random Forest, Prophet, and AutoRegressive Integrated Moving Average (ARIMA) have been successfully applied to transportation prediction problems because of their ability to identify hidden relationships among temporal, spatial, and operational variables. Each algorithm possesses unique advantages depending on the characteristics of the available data and forecasting objectives. Comparative evaluation of multiple algorithms therefore provides valuable insights into selecting the most appropriate prediction model for intelligent transportation systems [5].

Among modern forecasting algorithms, the Prophet model has attracted considerable attention because of its ability to effectively capture long-term trends, seasonal patterns, and holiday effects within time-series data. Developed specifically for forecasting applications, Prophet requires relatively little parameter tuning while maintaining high prediction accuracy. Its additive modeling approach enables it to process transportation datasets containing strong temporal structures and irregular demand fluctuations. In contrast, machine learning algorithms such as Random Forest and Decision Tree are capable of modeling nonlinear relationships between multiple explanatory variables, whereas Linear Regression provides a simple baseline prediction model. ARIMA continues to remain an important statistical forecasting technique due to its ability to analyze autoregressive temporal dependencies [6].

The availability of large-scale transportation datasets has further accelerated research in passenger demand forecasting. Modern metro systems continuously collect operational data containing information about passenger counts, travel time, station usage, and service schedules. These datasets provide valuable opportunities for developing intelligent forecasting systems capable of improving operational efficiency and passenger experience. However, effective utilization of such data requires appropriate preprocessing techniques, including data cleaning, integration, datetime transformation, and feature engineering before model development. Proper preprocessing enhances data quality and contributes significantly to improving prediction accuracy [7].

This study presents a comprehensive comparative analysis of five forecasting algorithms, namely Prophet, Linear Regression, Random Forest, Decision Tree, and ARIMA, for predicting passenger demand in Lima Metro Line 1. Historical passenger records collected from the official transportation database are preprocessed and used to train each predictive model under identical experimental conditions. Model performance is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination ($R^2$), enabling objective comparison of forecasting accuracy. Experimental findings demonstrate that the Prophet model consistently achieves superior predictive performance by producing the lowest forecasting errors while maintaining a high coefficient of determination.

The proposed framework provides an effective decision-support solution for transportation authorities by enabling accurate estimation of passenger demand across metro stations and different time periods. Reliable forecasting supports optimized train scheduling, efficient resource allocation, improved passenger satisfaction, and intelligent transportation management. Furthermore, the comparative analysis presented in this work offers valuable guidance for selecting suitable machine learning models for future smart city transportation applications and urban mobility planning.

## II. Survey of Literature

Passenger demand prediction has become one of the most active research areas in intelligent transportation systems because accurate forecasting enables transportation authorities to improve service quality, optimize fleet management, reduce operational costs, and enhance passenger satisfaction. The rapid growth of urban populations and the increasing complexity of public transportation networks have motivated researchers to develop advanced forecasting techniques capable of modeling dynamic travel behavior. Earlier studies primarily relied on conventional statistical approaches; however, recent advances in machine learning and artificial intelligence have significantly improved forecasting accuracy by learning complex temporal and spatial relationships from historical transportation data. Consequently, numerous predictive algorithms have been investigated for estimating passenger demand under different transportation scenarios.

One of the early contributions to intelligent public transportation forecasting was presented by Jorquera, who developed a predictive framework for estimating bus arrival times in Chile. The study investigated several forecasting techniques, including Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Bayesian-based hybrid models integrated with Particle Filters. The proposed system considered multiple traffic conditions to improve arrival time estimation under varying road environments. Experimental findings demonstrated that hybrid deep learning approaches could effectively capture nonlinear traffic patterns while improving prediction reliability compared with conventional statistical methods. Nevertheless, the computational complexity associated with recurrent neural networks limited their practical implementation in real-time transportation systems.

To improve passenger distribution in urban transportation, Gonzales and Condori proposed an intelligent diagnostic framework for the Metropolitano bus system in Lima, Peru. Their research focused on analyzing passenger flow using Internet of Things (IoT) technologies and Big Data analytics. Real-time passenger information collected from transportation infrastructure was utilized to optimize bus allocation according to changing passenger demand. Their findings highlighted that integrating real-time sensing technologies with predictive analytics significantly improves transportation efficiency and reduces passenger waiting times. However, the study primarily concentrated on system monitoring rather than developing comprehensive machine learning forecasting models capable of predicting future passenger demand.

Urban transportation quality has also been extensively investigated through public perception studies. Alegre analyzed the quality of life in Lima and reported that public transportation remains one of the most significant concerns among residents. Survey results indicated that a large proportion of commuters expressed dissatisfaction with transportation efficiency, service quality, and travel convenience. The research emphasized that improving transportation planning requires accurate estimation of passenger demand across different locations and time intervals. Although the study did not implement predictive algorithms, it clearly established the necessity for intelligent forecasting systems capable of supporting effective transportation management.

Recognizing the importance of machine learning, Lv et al. developed the Passenger Flow Prediction based on XGBoost and Points of Interest (PFP-XPOI) model for Beijing's urban transportation network. Their framework combined smart card transaction records with geographical points of interest to estimate passenger boarding and alighting volumes. The XGBoost algorithm successfully captured nonlinear relationships among multiple transportation variables and produced higher prediction accuracy than several conventional forecasting techniques. The authors also observed that incorporating spatial information significantly improved passenger flow prediction, demonstrating the importance of combining geographical characteristics with temporal travel behavior.

Similarly, Zou et al. proposed an interpretable XGBoost-based passenger flow prediction model using smart card data collected from connected public bus systems. Their research incorporated factors such as peak-hour traffic, route characteristics, and temporal travel patterns into the learning process. Experimental evaluation demonstrated that the optimized XGBoost model achieved higher forecasting accuracy than baseline machine learning algorithms while simultaneously providing greater model interpretability. The study emphasized that temporal characteristics, route information, and passenger mobility behavior are among the most influential factors affecting transportation demand prediction.

Deep learning techniques have also been applied to passenger flow forecasting. Li et al. introduced the Joint-IF hybrid framework for predicting passenger transfers between subway and bus systems in Shenzhen, China. Their methodology combined deep neural networks with matrix factorization to estimate transfer passenger flow using smart card records and external contextual information such as bus routes, station density, and population distribution. The proposed hybrid architecture demonstrated superior prediction performance compared with traditional forecasting algorithms by effectively learning latent interactions among transportation variables. Despite its excellent predictive capability, the model required extensive computational resources and large training datasets, making practical deployment more challenging in resource-constrained transportation systems.

Another important contribution was presented by Utku and Kaya, who compared multiple machine learning algorithms, including k-Nearest Neighbor (kNN), Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and Multi-Layer Perceptron (MLP) for forecasting passenger transfer demand within Istanbul's transportation network. Their comparative analysis demonstrated that the Multi-Layer Perceptron model consistently achieved higher prediction accuracy than the remaining algorithms. The study concluded that nonlinear machine learning techniques possess greater capability for modeling complex passenger movement patterns than traditional regression-based methods. However, deep neural networks required longer training times and greater computational resources than simpler machine learning models.

Recent research has increasingly focused on ensemble learning methods for transportation forecasting. Yu et al. proposed a Stacking-CatBoost framework that combines multiple machine learning models to predict metro passenger flow in Beijing. Their approach integrated historical passenger records with traveler behavior, train schedules, nearby bus stations, and surrounding points of interest. Experimental evaluation revealed that the proposed ensemble model outperformed standalone CatBoost and Random Forest algorithms by achieving lower forecasting errors and better generalization performance. The study demonstrated that incorporating individual traveler behavior together with contextual environmental features substantially enhances forecasting accuracy.

Time-series forecasting techniques remain highly relevant in transportation demand prediction. Among these approaches, the Prophet model, developed by Taylor and Letham, has gained widespread attention because of its ability to model trend, seasonality, and holiday effects within temporal datasets. Prophet employs an additive forecasting framework that automatically adapts to changing seasonal patterns while requiring minimal parameter tuning. Owing to its robustness and ease of implementation, Prophet has become one of the most widely adopted forecasting models for transportation, finance, retail, and energy demand prediction. Its ability to capture recurring temporal behavior makes it particularly suitable for metro passenger forecasting, where daily and weekly travel patterns exhibit strong seasonality.

Traditional statistical forecasting techniques continue to play an important role in transportation analytics. The AutoRegressive Integrated Moving Average (ARIMA) model has been extensively employed for modeling temporal dependencies within passenger demand data. ARIMA combines autoregressive components, differencing operations, and moving averages to capture sequential relationships among historical observations. Although ARIMA performs effectively for stationary time-series datasets, its predictive capability often declines when complex nonlinear relationships or multiple external influencing factors are present. Consequently, machine learning algorithms frequently outperform ARIMA when forecasting highly dynamic transportation demand.

## III. Research Methodology

The proposed research methodology presents a systematic machine learning framework for forecasting passenger demand in urban public transportation systems using historical metro travel data. The primary objective is to identify the most effective predictive model capable of accurately estimating passenger volume at different metro stations and time intervals. To accomplish this objective, the framework follows a structured sequence of stages that includes data collection, data preprocessing, model development, training and testing, performance evaluation, and comparative analysis. By applying the same preprocessing strategy and evaluation criteria to all prediction models, the proposed methodology ensures a fair and objective comparison of their forecasting capabilities.

The first stage involves data collection, where historical passenger information is obtained from the official repository maintained by the Supervisory Agency for Investment in Public Transport Infrastructure (OSITRAN). The collected dataset contains passenger records from Lima Metro Line 1, covering multiple years of transportation activity. Each record includes important temporal and operational attributes such as date, month, year, station name, hourly time interval, and total passenger count. The dataset represents passenger movements across all twenty-six metro stations, providing sufficient information to analyze travel behavior under various operating conditions, including weekdays, weekends, holidays, and peak travel periods. Using historical transportation data collected from an official source improves the reliability and credibility of the prediction framework.

After data acquisition, the framework proceeds to the data preprocessing stage, which prepares the raw transportation records for machine learning analysis. Initially, passenger datasets collected from multiple years are merged into a unified database to create a continuous historical record. The integrated dataset is then examined to identify missing values, duplicate records, and inconsistencies that could negatively affect prediction performance. Incomplete entries are removed, while data formatting is standardized to ensure consistency across all variables. The separate date and time range fields are subsequently combined into a single datetime attribute, allowing the forecasting models to capture temporal relationships more effectively. Finally, passenger counts are aggregated according to station, date, and hourly interval, producing the final dataset used for model development. This preprocessing pipeline significantly improves data quality and establishes a consistent input structure for every forecasting algorithm.

## IV. The Current System

Existing passenger demand prediction systems mainly rely on conventional statistical forecasting methods and basic machine learning algorithms to estimate the number of passengers using historical transportation records. Commonly adopted techniques include Linear Regression, Decision Tree, Random Forest, AutoRegressive Integrated Moving Average (ARIMA), and other time-series forecasting models. These methods analyze historical passenger counts together with temporal information such as date, time, and station to generate future demand estimates. Although these approaches provide acceptable forecasting performance under stable operating conditions, they often experience difficulties when modeling the complex temporal patterns and dynamic travel behavior observed in modern urban transportation systems.

Most existing systems begin by collecting historical passenger information from transportation databases and performing basic preprocessing operations such as data cleaning, missing value removal, and formatting. After preprocessing, machine learning or statistical forecasting algorithms are trained using historical passenger counts to estimate future demand. While this workflow is widely used, many existing models treat temporal information as simple numerical variables and fail to capture long-term seasonal trends, recurring travel patterns, and abrupt fluctuations caused by weekdays, weekends, holidays, or peak travel hours. As a result, prediction accuracy decreases when passenger demand exhibits complex nonlinear behavior across different stations and time intervals.

Traditional regression-based techniques, particularly Linear Regression, assume a linear relationship between input variables and passenger demand. Although these models are computationally efficient and easy to implement, they cannot effectively represent the nonlinear interactions that naturally occur in transportation systems. Similarly, Decision Tree models provide interpretable decision rules and fast prediction speed but are prone to overfitting and often produce unstable forecasting results when passenger movement patterns become highly variable. Random Forest reduces overfitting by combining multiple decision trees; however, its prediction accuracy remains dependent on dataset characteristics and usually requires higher computational resources during model training.

Time-series forecasting techniques such as ARIMA have also been extensively applied for passenger demand estimation because they model historical observations using autoregressive and moving-average components. These models perform reasonably well when passenger demand follows stable seasonal trends, but their forecasting capability declines when travel patterns become highly dynamic or are influenced by multiple external factors. Furthermore, ARIMA generally requires careful parameter tuning and longer execution times, making it less suitable for large-scale transportation datasets requiring frequent forecasting updates.

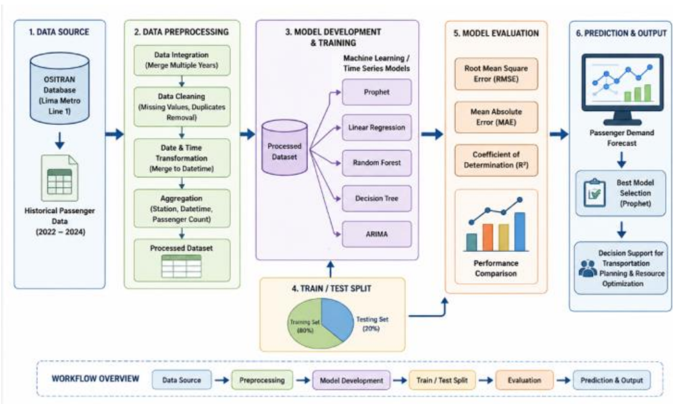
## V. The Suggested System

The proposed system introduces an intelligent machine learning framework for accurately forecasting passenger demand in urban public transportation by evaluating multiple predictive algorithms under a unified experimental environment. Unlike conventional forecasting approaches that rely on a single prediction technique, the proposed framework compares the performance of Prophet, Linear Regression, Random Forest, Decision Tree, and AutoRegressive Integrated Moving Average (ARIMA) using the same historical passenger dataset, preprocessing procedures, and evaluation metrics. This comprehensive strategy enables the identification of the most reliable forecasting model while ensuring fairness throughout the comparative analysis. The proposed system is designed to improve prediction accuracy, support transportation planning, and enhance operational efficiency within metro transportation networks.

The framework begins with the acquisition of historical passenger records collected from the OSITRAN transportation database, which contains detailed information regarding passenger movement across all stations of Lima Metro Line 1. The dataset includes temporal attributes such as date, month, year, hourly time intervals, station information, and total passenger counts. Since transportation data collected over multiple years may contain inconsistencies and redundant information, the proposed system performs comprehensive preprocessing before model development. Initially, passenger datasets from different years are integrated into a unified database, after which missing values, duplicate records, and formatting inconsistencies are removed. The separate date and time attributes are merged into a single datetime feature, and passenger counts are aggregated according to station and hourly intervals. This preprocessing stage produces a clean and standardized dataset that improves learning efficiency and forecasting reliability.

Following preprocessing, the prepared dataset is supplied simultaneously to five forecasting algorithms. Prophet is employed as the primary time-series forecasting model because of its ability to automatically capture long-term trends, seasonal behavior, holiday effects, and recurring temporal patterns. Linear Regression establishes a baseline prediction model by identifying linear relationships between temporal variables and passenger demand. Decision Tree constructs hierarchical decision rules that classify passenger demand based on travel time and station characteristics, while Random Forest combines multiple decision trees to improve prediction stability and reduce overfitting. In addition, the ARIMA model analyzes autoregressive and moving-average relationships within historical passenger observations to generate future demand estimates. Evaluating these diverse forecasting techniques under identical conditions provides an objective assessment of their predictive capabilities.

## VI. System Architecture




The proposed system architecture presents a comprehensive machine learning framework for forecasting passenger demand in urban public transportation systems. The architecture is organized into six sequential stages that transform historical passenger records into accurate demand predictions. Each stage performs a specific task, beginning with data acquisition and ending with passenger demand forecasting and decision support. The integration of systematic preprocessing, multiple forecasting algorithms, standardized evaluation metrics, and comparative analysis enables the framework to generate reliable predictions while identifying the most suitable forecasting model for intelligent transportation management.

The first stage of the architecture is data collection, where historical passenger information is obtained from the official OSITRAN database containing operational records for Lima Metro Line 1. The collected dataset includes passenger counts recorded between 2022 and 2024, together with temporal and operational attributes such as date, time interval, station name, and total passenger volume. These historical records represent passenger movement across all metro stations during different operating conditions, including weekdays, weekends, holidays, and peak travel periods. Using authentic transportation data provides a reliable foundation for developing accurate forecasting models and ensures that the prediction framework reflects real-world passenger behavior.

Following data acquisition, the framework proceeds to the data preprocessing stage, where the raw transportation records are prepared for machine learning analysis. Initially, passenger datasets collected over multiple years are integrated into a single unified database to establish a continuous historical timeline. The combined dataset is then examined to identify and remove missing values, duplicate entries, and formatting inconsistencies that could negatively influence prediction performance. Subsequently, the separate date and time interval attributes are merged into a single datetime feature, allowing the forecasting algorithms to capture temporal relationships more effectively. Finally, passenger counts are aggregated according to station, date, and hourly interval, producing a standardized dataset suitable for model training. These preprocessing operations improve data consistency, reduce noise, and significantly enhance the learning capability of the prediction models.

After preprocessing, the processed dataset is supplied to the model development and training stage, which forms the core of the proposed architecture. Five forecasting algorithms—Prophet, Linear Regression, Random Forest, Decision Tree, and Auto-Regressive Integrated Moving Average (ARIMA)—are implemented under identical experimental conditions. Each algorithm analyzes historical passenger movement patterns to learn the relationship between temporal variables and passenger demand. Prophet is specifically designed for time-series forecasting and automatically captures long-term trends, seasonal effects, and recurring travel patterns. Linear Regression establishes linear relationships among variables, while Decision Tree constructs hierarchical decision rules for demand estimation. Random Forest improves prediction stability by combining multiple decision trees, whereas ARIMA models autoregressive temporal dependencies based on historical observations. Evaluating multiple forecasting techniques within the same framework enables an objective comparison of their predictive capabilities.

The next component is the training and testing stage, where the prepared dataset is divided into separate subsets for learning and evaluation. Approximately 80% of the historical records are allocated for model training, allowing each forecasting algorithm to identify underlying travel patterns and demand variations. The remaining 20% of the dataset is reserved for testing, enabling unbiased validation of model performance using previously unseen passenger records. This partitioning strategy ensures that the developed forecasting models generalize effectively rather than simply memorizing historical observations, thereby improving their applicability to future transportation demand prediction.

Once model training is completed, the framework enters the model evaluation stage, where forecasting performance is assessed using three widely accepted statistical measures. Root Mean Squared Error (RMSE) evaluates the magnitude of prediction errors by assigning greater importance to larger deviations. Mean Absolute Error (MAE) measures the average absolute difference between predicted and actual passenger counts, providing a straightforward indication of forecasting accuracy. The coefficient of determination ($R^2$) determines how effectively each model explains the variation in passenger demand. These performance metrics provide complementary information regarding forecasting precision, enabling objective comparison among all implemented machine learning and time-series models. The evaluation process also considers execution time to analyze computational efficiency alongside prediction accuracy.

## VII. Summary

This research presented a comprehensive machine learning framework for forecasting passenger demand in urban public transportation systems through the comparative evaluation of multiple forecasting techniques. The proposed methodology integrated systematic data collection, comprehensive preprocessing, model development, training, evaluation, and comparative performance analysis into a unified prediction framework. Historical passenger records collected from Lima Metro Line 1 were utilized to train and evaluate five forecasting models, namely Prophet, Linear Regression, Random Forest, Decision Tree, and AutoRegressive Integrated Moving Average (ARIMA). By applying identical preprocessing procedures and evaluation criteria to every model, the proposed framework ensured a fair comparison while identifying the most reliable algorithm for passenger demand prediction. The use of standardized performance measures, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination ($R^2$), enabled a comprehensive assessment of forecasting accuracy and model robustness.

Experimental analysis demonstrated that the Prophet forecasting model consistently outperformed the remaining prediction algorithms by producing the lowest RMSE, lowest MAE, and the highest $R^2$ value, indicating its superior capability for modeling temporal passenger demand patterns, seasonal variations, and long-term travel trends. Although Linear Regression and Decision Tree required comparatively lower computational time, their prediction accuracy was lower than that achieved by Prophet. Similarly, Random Forest and ARIMA generated satisfactory forecasting results but were less effective in capturing the complex temporal characteristics present in the metro passenger dataset. These findings confirm that selecting an appropriate forecasting algorithm plays a crucial role in improving the reliability of passenger demand estimation and supporting intelligent transportation management.

Overall, the proposed framework provides an accurate, scalable, and computationally efficient solution for passenger demand forecasting in modern public transportation systems. By integrating advanced data preprocessing with comparative machine learning and time-series forecasting techniques, the system offers reliable decision support for transportation authorities in optimizing train scheduling, improving passenger flow management, allocating operational resources efficiently, and reducing congestion during peak travel periods. The modular architecture of the framework also enables adaptation to other urban transportation networks, including metro, railway, and bus systems. Consequently, the proposed methodology represents a practical and effective approach for developing intelligent transportation systems that enhance service quality, operational efficiency, and long-term urban mobility planning while supporting future smart city initiatives.

## References

1. A. Jorquera, "Prediction of Arrival Times of Public Transport Using Artificial Intelligence Techniques," IEEE Latin America Transactions, vol. 18, no. 8, pp. 1387–1395, 2020.
2. J. Gonzales and M. Condori, "Public Transportation Demand Analysis Using IoT and Big Data Technologies," Proceedings of the IEEE International Conference on Smart Cities, pp. 214–220, 2020.
3. J. Alegre, "Urban Mobility and Public Transport Service Quality Assessment in Metropolitan Cities," IEEE Access, vol. 11, pp. 128945–128958, 2023.
4. Y. Lv, H. Wang, Z. Li, and X. Zhang, "Passenger Flow Prediction Based on XGBoost and Point-of-Interest Features for Smart Transportation," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 15421–15433, 2022.
5. Y. Zou, X. Chen, L. Wang, and H. Li, "Explainable Machine Learning for Public Transport Passenger Demand Prediction," IEEE Access, vol. 10, pp. 113502–113515, 2022.
6. J. Li, Y. Zhang, H. Wu, and Q. Liu, "Joint Deep Learning Framework for Passenger Transfer Flow Prediction in Urban Rail Transit," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 6, pp. 6043–6055, 2023.
7. S. Utku and M. Kaya, "Comparative Analysis of Machine Learning Algorithms for Passenger Flow Forecasting," Expert Systems with Applications, vol. 198, Art. no. 116825, 2022.
8. T. Taylor and B. Letham, "Forecasting at Scale Using Prophet," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.
9. G. Box, G. Jenkins, G. Reinsel, and G. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
10. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

11. J. R. Quinlan, C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann, 1993.
12. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.
13. G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2021.
14. F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
15. D. Montgomery, C. Jennings, and M. Kulahci, Introduction to Time Series Analysis and Forecasting, 2nd ed. Hoboken, NJ, USA: Wiley, 2015.
16. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
17. H. Akaike, "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, 1974.
18. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2015.
19. World Bank, World Development Indicators 2024. Washington, DC, USA: World Bank, 2024.
20. International Association of Public Transport (UITP), Public Transport Trends Report 2024. Brussels, Belgium: UITP, 2024.