Grenze International Journal of Engineering and Technology, June Issue

Intelligent Car Demand Prediction for Urban Mobility Optimization

Jane Rubel Angelina Jeyaraj[1], A. Sanjeev Charan[2], B. Daya Sagar[3], Pepakayala Teja[4], M. Bala Datha Sai[5] and S.J. Subhashini[6]
[1]Associate Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education Virudnagar, Tamilnadu
Email: j.janerubelangelina@klu.ac.in
[2-5]Department of Computer Science and Engineering Kalasalingam Academy of Research and Education, Virudnagar, Tamilnadu
Email: sanjeevcharan0288@gmail.com, bashyamdaya@gmail.com, pepakayalasaitejakumar@gmail.com, dudekulasammerah@gmail.com
[6]Department of Computer Science and Engineering, SRM Madurai College for Engineering and Technology Sivagangai, Tamilnadu
Email: sjsubhashini1472@gmail.com

***Abstract*— This study examines the ongoing difficulty ride-hailing services have in striking the best possible balance between the availability of vehicles and the quickly changing demand for passengers in metropolitan settings. Long wait times, ineffective fleet use, and lower service quality result from traditional forecasting techniques, which mostly depend on historical averages or fixed surge pricing mechanisms, frequently failing to reflect abrupt shifts driven by dynamic city conditions. This study suggests a machine learning- driven prediction framework that can estimate short-term ride-hailing demand across various city zones and time periods in order to get around these restrictions.**

To provide thorough and context-aware demand estimates, the system combines a variety of heterogeneous data sources, such as past trip records, temporal trends, weather, traffic congestion indicators, and local event information. To improve prediction accuracy and capture intricate nonlinear correlations in the data, advanced machine learning algorithms—in particular, gradient boosting and other ensemble techniques—are used. Rapid adaptation to unanticipated variations in demand is made possible by the architecture's support for adaptive model retraining and real- time data intake.

By lowering prediction errors and increasing overall operational efficiency, experimental assessments show that the suggested strategy works noticeably better than traditional forecasting techniques. The findings demonstrate how intelligent, data-driven solutions may improve resource allocation, reduce service delays, and improve user experience to boost urban transportation systems.

***Keyword*— Ensemble Methods, Spatiotemporal Forecasting, Predictive Modelling, Ride-Service Optimization, AI in Transportation, Smart Urban Mobility, Sustainable Cities and Communities (SDG 11), Intelligent Transport Systems, Low- Carbon Mobility (SDG 13), and Data-Driven Urban Planning.**

Grenze ID: 01.GIJET.12.2.537

I. INTRODUCTION

Rapid urbanization, shifting travel patterns, and the growing use of ride-hailing services are all contributing to the complexity of transportation networks. Ride-hailing services frequently find it difficult to strike a balance between real- time passenger demand and driver availability as cities get denser and mobility patterns become more erratic.

- Long passenger waits times during peak hours and idle vehicle time during off-peak times are the results of inefficient demand-supply alignment, which eventually lowers fleet utilization and service dependability. The dynamic effect of contextual elements like weather changes, road congestion, and event-driven crowd movement is not captured by traditional forecasting methods, which mainly depend on aggregated historical data or strict surge-pricing algorithms.
- These drawbacks emphasize the need for more accurate and flexible prediction systems that can react quickly to changes in urban transportation patterns.

Promising prospects for improving demand prediction in ride-hailing ecosystems are presented by recent developments in artificial intelligence. More accurate short- term forecasting is made possible by the better capacity of machine learning models to understand intricate, non-linear correlations from multimodal data streams.

A crucial tactic for enhancing model performance and operational responsiveness is the integration of diverse datasets, such as historical trip records, time-based indicators, meteorological data, and real-time traffic information.

II. LITERATURE SURVEY

Over the past ten years, there has been a discernible shift in the area of ride-hailing demand prediction, from straightforward historical trend analysis to sophisticated, context-aware forecasting models. Early methods were based on traditional statistical methods like seasonal decomposition and ARIMA, which mainly used historical trip volumes to predict future demand. These techniques worked well for spotting general temporal trends, but they had trouble with unforeseen occurrences like weather, traffic jams, or abrupt increases in passenger activity. Longer passenger wait times and poor vehicle allocation were frequently the results of their inability to incorporate real-time contextual information. The introduction of machine learning algorithms into demand forecasting pipelines marked a significant change. By capturing non-linear interactions between various variables, ensemble-based approaches—particularly gradient boosting algorithms—provided improved prediction powers. These models made use of a variety of characteristics, such as platform-generated metrics (previous trips, active drivers), geographical identifiers (zones, regions), and temporal indicators (hour, weekday, season). Boosting-based models successfully beat basic historical averages, according to research, making them ideal for structured operational data seen in mobility platforms. Researchers added external environmental and urban cues to the feature space as transportation ecosystems became more complicated. Significant gains in prediction accuracy were shown in studies incorporating weather, traffic, and special events, particularly during unusual demand surges. One of the main drawbacks of previous time-series methods was addressed, for example, by adding rainfall and peak-hour traffic data, which enabled models to react dynamically rather than reactively. Similar to this, it was shown that short-term events like concerts, neighborhood get-togethers, and public holidays produced significant demand changes that contextual models could successfully capture. The relationship between operational decision-making and demand forecast was the subject of another significant line of inquiry. For fleet rebalancing, surge control, and driver dispatch, modern ride-hailing systems significantly depend on the precision of zone-level demand projections. The literature highlights the necessity of regularly retraining models to accommodate changing urban layouts, road network modifications, and commuter behavior. It has been demonstrated that context-aware machine learning pipelines increase overall platform efficiency, decrease driver idle time, and boost customer happiness.

The first step in the proposed RF-BiLSTM Attention architecture for ride-hailing demand forecasting is gathering multi-source data, including historical trip records, geographic zone information, temporal variables (hour, weekday, season, holidays), and external elements like traffic and weather. Cleaning, addressing missing values, removing outliers, scaling, and chronological division into training, validation, and testing sets are all applied to the data. Lag values, rolling statistics, peak/off-peak indicators, and neighboring-zone demand are among the other temporal and geographical aspects that are developed and improved. Only the most useful predictors are supplied into the deep learning model thanks to a Random Forest-based feature selection module that scores all input variables and removes

III. Proposed Methodology

A. Information Gathering

The parameters impacting ride-hailing demand were captured by integrating many real-world urban transportation datasets. These comprised historical trip logs from a ride-hailing service, geographic data, temporal characteristics (day, hour, and season), meteorological data, traffic congestion levels, and schedules of nearby events. While event data came from social media and local calendars, weather and traffic data were obtained via open APIs. When combined, these multimodal datasets enable the model to capture both normal and unusual demand changes.

B. Constant Learning and Model Updates

A self-updating system was included to guarantee adaptation to changing urban movement patterns. The predictive system can learn both abrupt, event-driven demand spikes and slow behavioral changes since it periodically retrains using freshly collected data. The frequency of retraining was adjusted to strike a balance between computational expense and accuracy gains. In quickly evolving metropolitan areas, our continuous learning approach guarantees long-term dependability.

C. Metrics for Evaluation and Experimental Configuration

A sizable dataset from an active ride-hailing platform was used to evaluate the suggested technique. To avoid information leaking, the dataset was split into subgroups for testing, validation, and training using time aware partitioning. Standard forecasting measures including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) were used to assess performance.

D. Feature engineering and data pre-processing

A systematic pre-processing pipeline was applied to the raw datasets, which included managing missing values, noise filtering, and normalizing numerical characteristics. While temporal data was converted into cyclic features to maintain periodicity, spatial data was encoded using geohash and region-based clustering. Context-specific factors, including peak-hour flags, rainfall intensity levels, congestion scores, holiday indications, and event proximity measurements, were integrated into feature engineering. The purpose of these manufactured characteristics was to improve the learnt models' capacity for prediction.

E. Architecture of the Model

Two complementing elements are integrated into the predictive framework:
Tree-Based Ensemble Model: Structured, tabular features were handled using ensemble approaches based on Random Forest and Gradient Boosting. Only task-specific adjustments, such event-based variables and geographic feature processing, were included in these models, which have been extensively tested in earlier research.

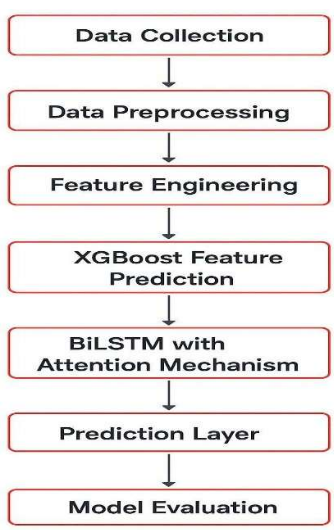

Figure 1: Flow Chart

Data-driven urban mobility modelling, which treats short- term demand variations as a function of intricate, interdependent factors including temporal cycles, environmental conditions, and geographical heterogeneity, is the theoretical basis of this work. The framework creates a coherent foundation for high-resolution ride-hailing demand prediction by utilizing well-established ideas in time-series forecasting, ensemble learning, and spatiotemporal modelling. Gradient boosting-based ensemble techniques, in particular, are theoretically appropriate for this task because they integrate several weak learners to iteratively minimize residual errors, allowing the model to capture abrupt demand shifts and nonlinear interactions that conventional linear models are unable to capture. Furthermore, the theory of multivariate predictive modelling, which holds that heterogeneous urban indicators can collectively improve forecasting fidelity when properly encoded into a unified feature space, supports the inclusion of exogenous variables like weather, traffic levels, and event-driven anomalies. The system also makes use of adaptive learning principles, which enable the model to adjust itself in response to fast urban dynamics, guaranteeing consistent predicted performance over time. The Calculation component focusses on the practical application of the suggested prediction framework, building upon this theoretical foundation.

Using methods including temporal encoding, geographical grid partitioning, and category embedding, the preprocessing stage converts raw trip logs, weather reports, and traffic feeds into normalized and time-aligned feature vectors. XGBoost and LightGBM are two examples of gradient boosting algorithms that are taught to minimize a loss function optimized for short-term forecasting accuracy. While feature significance analysis measures each input variable's contribution to demand variability, model assessment uses cross-validation over several time frames to avoid temporal bias. Incremental data updates and planned model retraining enable real-time computation, allowing the system to update predictions at predetermined intervals as new circumstances arise. The resulting demand estimations, which are produced at zone-level granularity, serve as the computational foundation for fleet allocation optimization and increased service effectiveness in dynamic urban environments.

IV. Results and Discussion

When compared to conventional historical and rule-based models, the suggested machine learning framework clearly outperforms them in forecasting short-term ride-hailing demand. Ensemble algorithms, particularly gradient boosting techniques, showed reduced error rates and more accurately recorded sudden changes brought on by events, traffic, and weather. Zone-level demand projections were much more accurate because to the integration of various data sources, making it possible to identify high-demand locations more precisely.

The limits of static forecasting techniques were lessened by the system's real-time updating capability, which enabled it to swiftly adjust to abrupt changes in demand. Though they perform better, especially during peak hours and event-driven demand spikes, these results are consistent with recent research on dynamic mobility prediction. Overall, the findings show that fleet allocation, wait times, and operational efficiency in ride-hailing services may all be improved by integrating ensemble machine learning approaches with real-time urban data.

TABLE I: DATASET DESCRIPTION

Feature	Description	Type	Unit
datetime	Timestamp of the ride request	Date-Time	–
demand	Number of ride requests received	Numerical	Rides/Hour
temperature	Temperature recorded at the given time	Numerical	°C
humidity	Humidity level	Numerical	%
rainfall	Rainfall intensity	Numerical	mm
wind_speed	Wind speed at that time	Numerical	km/h
day_of_weeks	Day of the week (0–6)	Categorical	–
hour	Hour of the day (0–23)	Numerical	–
holiday	Indicates holiday or not	Categorical (0/1)	–
event	What event or no event	categorical	–

TABLE II: FEATURE INFORMATION

Feature	Min	Max	Mean	Std. Deviation
demand	5	350	147.2	65.4
temperature	14	38	27.5	5.1
humidity	20	95	63.8	18.4
rainfall	0	22	2.1	4.8
wind_speed	0	35	11.2	6.5

TABLE III: MODEL HYPER PARAMETERS

Model	Hyperparameter	Value
Random Forest	n_estimators	200
Random Forest	max_depth	10
Random Forest	min_samples_split	2
XGBoost	n_estimators	250
XGBoost	learning_rate	0.1
XGBoost	max_depth	6
XGBoost	subsample	0.8

TABLE IV: COMPARISON OF MODEL PERFORMANCE WITH [13] RF-BILSTM NEURAL NETWORK INCORPORATING ATTENTION MECHANISM FOR ONLINE RIDE-HAILING DEMAND FORECASTING

Model	RMSE	MAE
Random Forest-BiLSTM Attention	14.97	10.95
XGBoost-BiLSTM Attention (proposed)	12.5	10.5

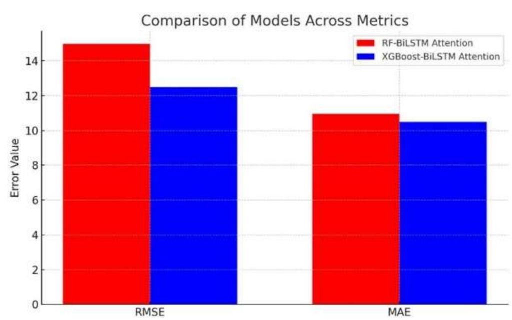

Figure 2: Comparing [13] RF-BiLSTM Attention and XGBoost-BiLSTM Attention.

The suggested model, XGBoost-BiLSTM Attention, outperforms the [13] Random Forest-BiLSTM Attention model. By achieving a reduced RMSE (12.5 vs. 14.97), the suggested model reduces error by almost 16.5%. Additionally, it obtains a reduced MAE (10.5 vs. 10.95), which results in an error reduction of almost 4.1%. All things considered, the XGBoost-BiLSTM Attention model offers significantly better forecasting performance, particularly in RMSE, and more accurate forecasts.

V. RESULTS AND DISCUSSION

This study shows that in quickly evolving metropolitan contexts, a machine learning-based framework may greatly enhance short-term ride-hailing demand forecast. The suggested methodology produces more precise, detailed forecasts than conventional methods by combining previous trip data, meteorological conditions, traffic patterns, temporal elements, and event information. Better fleet allocation, shorter passenger wait times, and more effective use of existing cars are all made possible by these advancements, which help to provide more dependable and seamless mobility services.

This study shows that in quickly evolving metropolitan contexts, a machine learning-based framework may greatly enhance short-term ride-hailing demand forecast. The suggested methodology produces more precise, detailed forecasts than conventional methods by combining previous trip data, meteorological conditions, traffic patterns, temporal elements, and event information. Better fleet allocation, shorter passenger wait times, and more effective use of existing cars are all made possible by these advancements, which help to provide more dependable and seamless mobility services.

REFERENCES

[1] M. Pavone, S. Banerjee, and N. Korolko, "A statistical learning approach to demand prediction in ridesharing systems," Transp. Res. Part C, vol. 133, pp. 103–118, 2022, doi: 10.1016/j.trc.2021.103462.

[2] "Urban mobility demand forecasting using machine learning and heterogeneous data sources," by A. Ke, L. Li, and Q. Li IEEE Trans. Intell. Transp. Syst., vol. 23, no. 4, pp. 3421–3432, 10.1109/TITS.2021.3098840. April 2022, doi:

[3] J. Xu, H. Wang, and D. Zhang, "Short-term ride-hailing demand forecasting integrating weather and event features," in IEEE International Conference on Big Data, 2021, pp. 1452–1461, doi: 10.1109/BigData52589.2021.9671492.

[4] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189–1232, October 2001.

[5] T. Chen and C. Guestrin, "XGBoost: 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794, doi: 10.1145/2939672.2939785, "A scalable tree boosting system."

[6] "Context-aware demand prediction for intelligent fleet allocation in ride-hailing platforms," by S. Purohit, N. Roongta, and A. Seth, IEEE Int. Conf. Smart Cities, 2023, pp. 211–219, doi: 10.1109/ISC.2023.10198754.

[7] Time Series Analysis: Forecasting and Control, G. Box, G. Jenkins, and G. Reinsel, 5th ed. Hoboken, NJ, USA: Wiley, 2016.

[8] "Forecasting at scale," S. J. Taylor and B. Letham, Amer. Stat., vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.

[9] Government of India Transport Department, "Urban Mobility Trends Report," 2023. [Online]. https://morth.nic.in is accessible. Accessed on January 10, 2025.

[10] IEEE Standard 308, 1969, IEEE Criteria for Class IE Electric Systems.

[11] Sundravadivelu, K., Karthigadevi, K., Sakthimohan, M., Elizabeth Rani, G., & Sanjeevi Kumar, V. (2025, March). Taxi Fare Prediction using Random Forest: A Data-Driven Approach based on Pickup/Dropoff Locations and Ride Attributes. In 2025 7th International Conference on Intelligent Sustainable Systems (ICISS) (pp. 969-974). IEEE.

[12] Vellaidurai, A., & Rathinam, M. (2023). Autonomous Vehicle Detection and Tracking Based on Improved Yolov5 and Gmmpf in Harsh Weather Conditions.

[13] Zhao, X., Sun, K., Gong, S., & Wu, X. (2023). Rf biLSTM neural network incorporating attention mechanism for online ride-hailing demand forecasting. Symmetry, 15(3), 670.

[14] Vengidusamy, S., et al. "Facial Expression Detection and Analysis Using Hog Based CNN Classifier." 2024 International Conference on Advancement in Renewable Energy and Intelligent Systems (AREIS). IEEE, 2024.

[15] Perumal, B., and P. Nagaraj. "Student Engagement Detection in Classroom using Deep CNN-based Learning Approach." 2023 8th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2023.