10.48047/jocaaa.2025.34.06.18

Intelligent Forecasting of City Bike-Share Trends for Sustainable Urban Transport

Mr P. Venkata Siva[1]|Mr G. Bharath Kumar[2]|Mr Sk. John Sydulu[3]|Sk.Hazi Hassain[4].

[1]Assistant Professor Department of CSE, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India
[2]Assistant Professor Department of CSE, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India
[3]Assistant Professor Department of CSE, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India
[4]PG Scholar Department of CSE, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

Abstract: Bike-sharing systems play a significant role in promoting sustainable urban mobility. However, their planning and operation present several challenges. One of the primary operational issues is the uneven distribution of bicycles across the service area, leading to shortages in some zones and surpluses in others. To maintain a satisfactory level of service, operators must conduct costly repositioning operations. These repositioning strategies typically depend on accurate predictions of expected bike pickups and returns at each station. Inaccurate forecasts can result in sub-optimal repositioning, reducing system efficiency. Therefore, developing reliable methods for forecasting bike usage is essential. This study addresses this issue by applying machine learning regression techniques to predict bike-sharing demand using historical usage and weather data. Three methods like Random Forest, Gradient Boosting are evaluated using data from the New York City bike-sharing system. The study outlines the model variables and the calibration process for each technique. Results are compared not only in terms of prediction accuracy but also in terms of the practical implementation of each method. Findings reveal that while all three techniques offer similar accuracy, the simpler calibration process of the Random Forest model makes it the most practical choice for most applications.

Key Words: Bike Rental, Shared-Mobility, Predictive Analytics, Linear Regression, Random Forest Regression, Gradient Boosting Regression, Demand Prediction.

1. Introduction

Bike-sharing systems have become a vital part of sustainable urban mobility, offering an environmentally friendly alternative to private vehicles and public transport. They help ease traffic congestion, reduce carbon emissions, and encourage healthier lifestyles. However, managing these systems efficiently remains a challenge due to the uneven distribution of bicycles across docking stations. This study examines the development and growing role of Seoul's bike-sharing program, launched in 2015 to address urban issues like congestion and air pollution. With daily rentals increasing from 65,000 in 2020 to 87,800 in 2021, the system

10.48047/jocaaa.2025.34.06.18

now plays a key role in both daily commuting and tourism. Nevertheless, fluctuating demand across different times and locations leads to operational difficulties. A major issue is the imbalance of bike availability—some stations run out of bikes while others are overloaded, making returns difficult. This necessitates frequent and costly repositioning efforts to maintain system balance and service quality.

To address these challenges, the study employs machine learning techniques such as OLS regression, Random Forest, and Gradient Boosting to analyze factors like temperature, time of day, season, and holidays for predicting demand patterns. The resulting insights can support more effective bike distribution strategies, enhancing the system's efficiency and overall user experience in Seoul's dynamic urban environment.

2. Literature survey

A Study on Bicycle-Sharing Dispatching Station Site Selection and Planning Based on Multivariate Data" by Lei, Zhang, and Ren, published in *Sustainability* in 2023, addresses the optimization of dispatching stations in bicycle-sharing systems. The authors focus on Xiamen Island, China, utilizing a Mean-shift clustering algorithm to group electronic fence parking points, thereby establishing dispatching stations and defining management areas. They further employ singular value decomposition to analyze travel demand patterns and characteristics. Additionally, regression models are constructed to explore the relationship between the urban built environment and bicycle-sharing trips during peak hours. The study identifies spatial and temporal differences in how the urban environment impacts bicycle-sharing usage, noting that residential and enterprise point-of-interest densities have opposite effects during morning and evening peaks. The findings aim to inform the planning of dispatch stations and the differentiated management of bicycle-sharing, ultimately reducing operational costs and promoting sustainable urban transportation.

3. Existing System

The concept of bike-sharing consists of providing a fleet of bicycles for users to use to make trips without needing to own them. This shifts the focus to a mobility-as-a-service model. From the user's perspective, bike-sharing, as with other vehicle-sharing initiatives (e.g., car-sharing, motor bike-sharing, scooter-sharing),offers the advantages of low-cost on-demand transportation(i.e., flexibility, traveling when and where needed)in front of the traditional public transportation alternatives with predefined routes and schedules

A predictive analytics approach for forecasting bike rental demand using logistic regression involves modelling the relationship between bike rental activity and various influencing

10.48047/jocaaa.2025.34.06.18

factors such as historical usage patterns, weather conditions, time of day, and day of the week. Logistic regression, although typically used for classification, can be adapted to predict the probability of high or low rental demand by categorizing usage levels. The process includes data pre-processing, feature selection, model training, and performance evaluation using metrics like accuracy, precision, and recall. While logistic regression is relatively simple and interpretable, its performance may be limited when capturing complex, non-linear relationships in the data compared to more advanced machine learning models. Nonetheless, it provides a fast, transparent baseline for understanding demand trends in bike-sharing systems.

4. Proposed System

The proposed system for forecasting bike rental demand using predictive analytics involves several key stages: data collection, preprocessing, model training, and prediction. First, historical bike rental data along with relevant external factors such as weather conditions, time, and location are gathered. The data is then cleaned and transformed to ensure quality and consistency.

i. Dataset and preprocess: In a predictive analytics approach for forecasting bike rental demand, the dataset typically includes historical bike rental records, weather data (temperature, humidity, wind speed, precipitation), and temporal information (hour, day, month, weekday/weekend, holidays). The preprocessing stage involves cleaning the data by handling missing values, removing outliers, and converting categorical variables into numerical formats using techniques like one-hot encoding. Feature scaling or normalization may also be applied to ensure model efficiency.

10.48047/jocaaa.2025.34.06.18

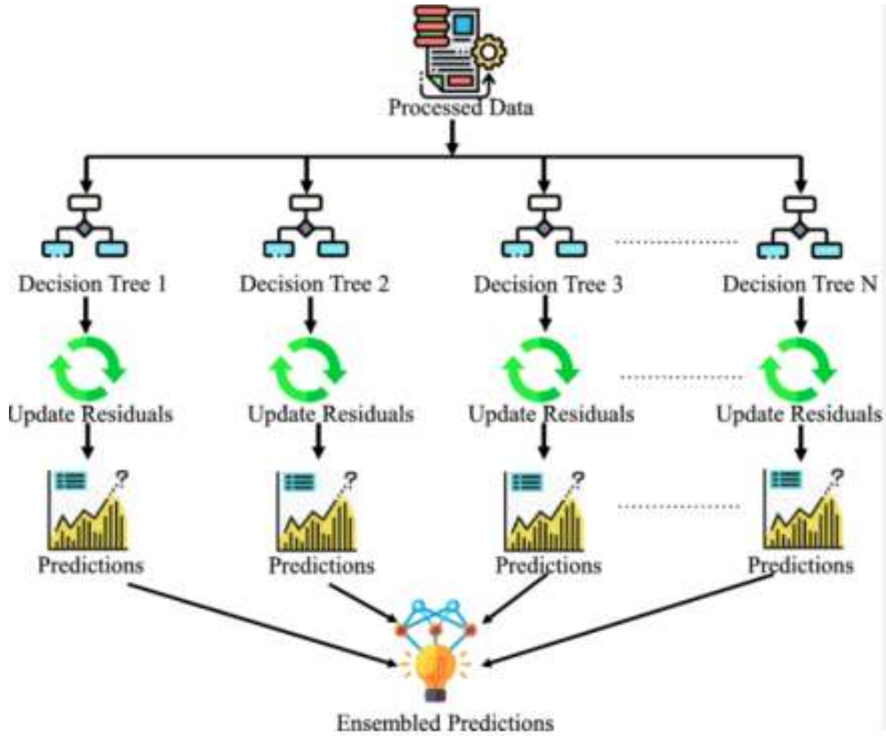

Fig 1: Predictive Analysis

ii. Model evaluation: The proposed system for forecasting bike rental demand using predictive analytics leverages advanced machine learning techniques, specifically Random Forest and Gradient Boosting. The process begins with the collection of historical rental data, weather conditions, and temporal variables such as time of day and day of the week. After thorough data preprocessing and feature engineering, both Random Forest and Gradient Boosting models are trained to learn complex patterns and relationships within the data. Random Forest offers robustness and ease of calibration, while Gradient Boosting provides higher predictive accuracy by sequentially correcting errors from previous models. The system then uses these trained models to forecast demand at various locations and times, enabling more efficient bike redistribution and improved operational planning in bike-sharing systems.

5. System Architecture

This architecture presents the data preparation and data pre-processing that were done in order to apply different machine-learning schemes, selection of learning schemes, selection of evaluation methods for the models and Implementation for data processing, transformations, analysis and machine learning.

10.48047/jocaaa.2025.34.06.18

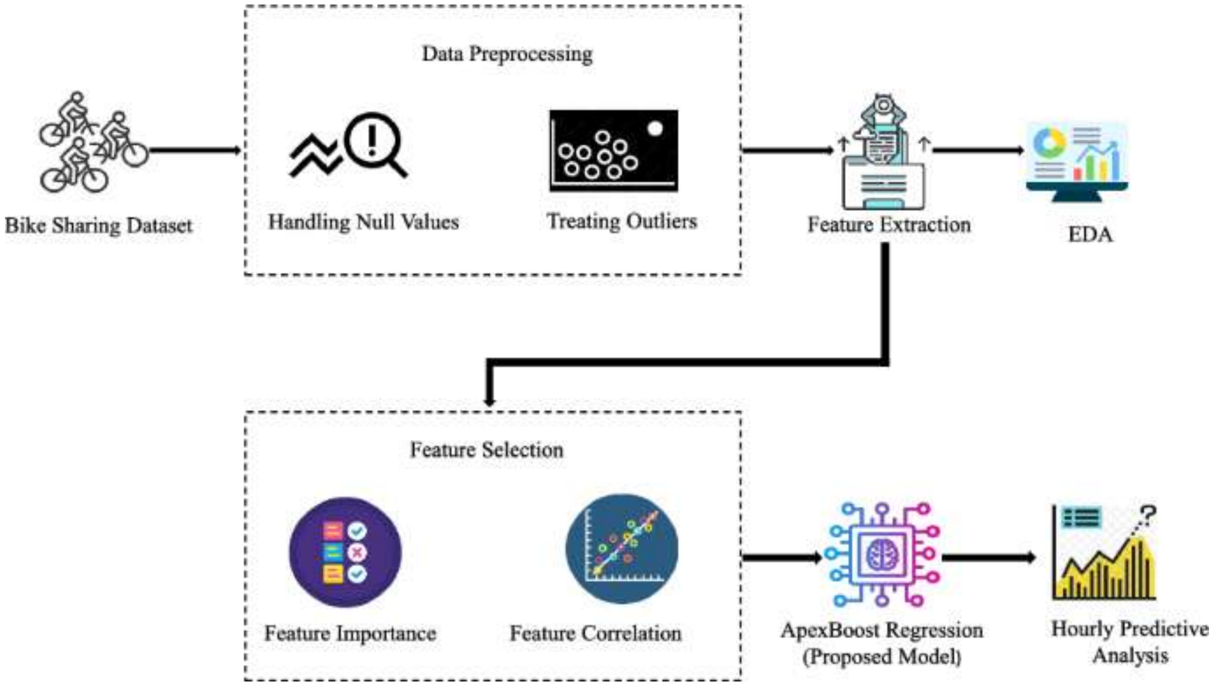

Fig 2: System Architecture

The developed prediction tool is based on a data-driven model that fuses information from three different sources: a static file with information and the location (coordinates) of bike sharing system stations, the system's rental records database, which includes spatial and temporal data for both the start and end of each rental, and weather data for the area where the system operates as well as a short-term weather forecast. For model training and prediction, three alternative time intervals are considered: one hour, two hours, and three hours.

6. Implementation Modules

The implementation of a predictive analytics approach for forecasting bike rental demand involves several key modules.

Data Collection Module gathers historical rental, weather, and temporal data.

Data Preprocessing Module cleans and transforms the data, handling missing values, encoding categorical variables, and generating relevant features.

Model Development Module applies machine learning algorithms such as Random Forest and Gradient Boosting, training them on the processed data.

Prediction Module uses the trained models to forecast demand across different locations and times.

Visualization and Decision Support Module presents the forecasts through dashboards or interfaces, helping operators optimize bike allocation and improve service efficiency.

10.48047/jocaaa.2025.34.06.18

7. Results and Discussion

This project forecasts city bike-sharing demand using historical data and predictive analytics to improve bike availability, rebalancing, and user experience. Model analysis shows that Linear, Ridge, and Lasso Regression struggle with complex demand patterns, with Lasso performing the worst. In contrast, Random Forest and Gradient Boosting offer superior accuracy by capturing non-linear and time-based trends. The results highlight the effectiveness of machine learning in understanding bike-share usage and the importance of adaptive forecasting systems for efficient urban mobility management.

Demand Patterns Identified

Working Days (Weekdays):Peak demand observed during morning (08:00-09:00) and evening (17:00-18:00) hours**,** matching commuting trends**.** Demand drops mid-day and late night, indicating lower bike usage outside of work hours. The model captures general trends well but underestimates peak demand.

Non-Working Days (Weekends & Holidays):A single peak observed between 10:00-14:00**,** showing leisure and recreational usage. Demand gradually declines post-afternoon**,** unlike the double peaks on weekdays. The model struggles to capture sudden fluctuations in late hours.

The image shows **a** bike-sharing dataset with 10,886 rows and 11 columns, containing hourly records of bike rentals along with weather and temporal factors. Here's a breakdown of the dataset and its relevance to forecasting city bike share trends shown below fig

```
print('Shape of data: ', mydata.shape)
```

```
Shape of data:  (10886, 11)
```

```
mydata.head(3)
```

datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	13	16
2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	32	40
2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	27	32

Fig 3: Demand Pattern

10.48047/jocaaa.2025.34.06.18

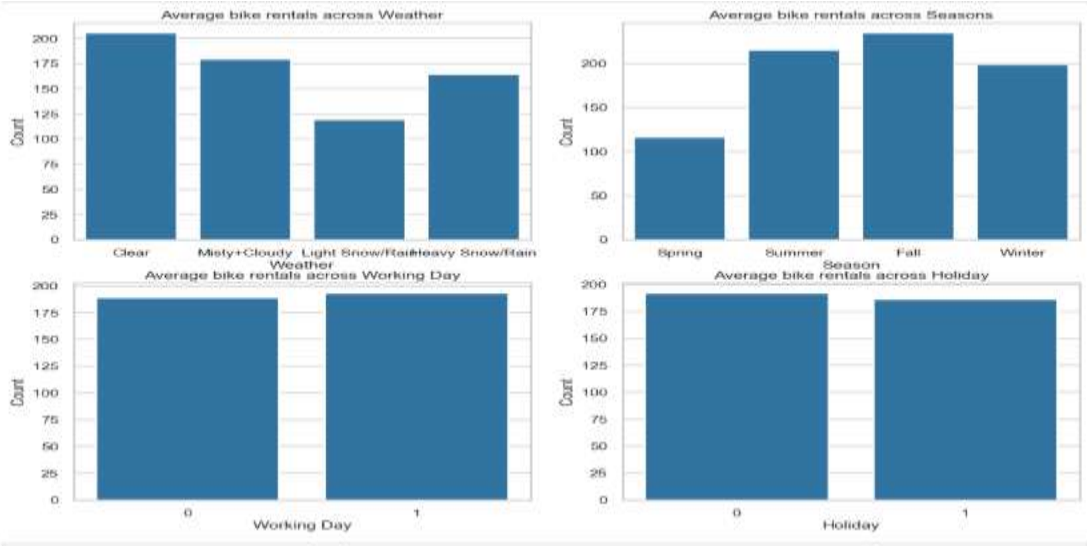

Fig 4: Overview Graph

The above bar charts show that bike rental demand is strongly influenced by weather and seasons. Rentals peak during clear weather and fall season, while they drop during light snow or rain and in spring. In contrast, there is little difference in rentals between working days and holidays, indicating that environmental conditions have a greater impact on bike usage than the day type.

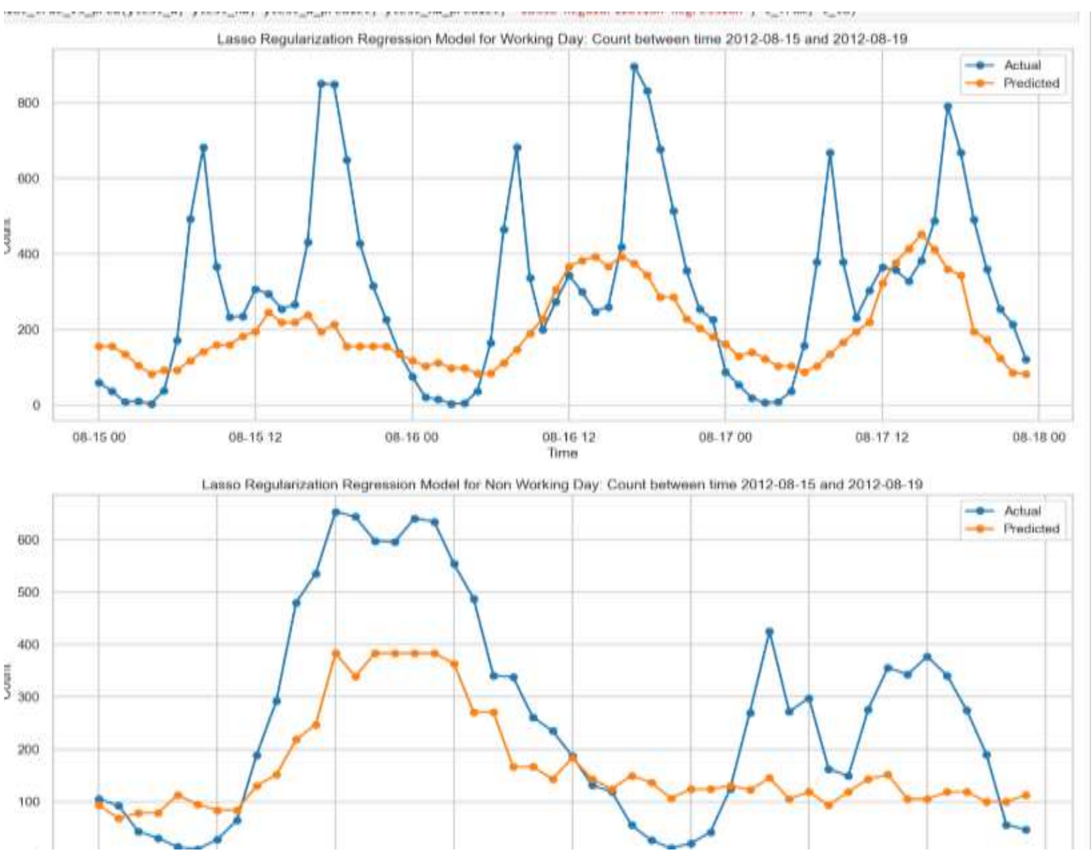

Fig 5: Work Day and Non Work Day

10.48047/jocaaa.2025.34.06.18

The Lasso model captures overall bike rental trends but underestimates peak demand during commute hours and mid-day leisure peaks on non-working days. It performs better during stable periods but struggles with sudden fluctuations, leading to overly smooth and less accurate predictions.

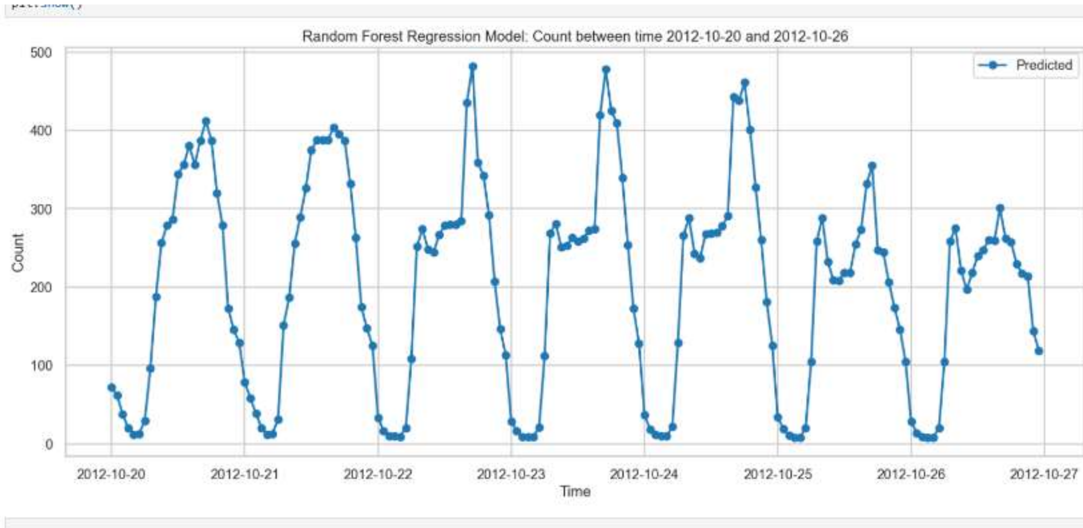

Fig 6: Random Forest Regression Model

The Random Forest model closely tracks daily bike-rental cycles from 20–26 Oct 2012, accurately reproducing the sharp morning and evening peaks and intervening lulls. Its ability to capture these repeating, non-linear patterns highlights its reliability for short-term demand forecasting.

odelling Algo	Train RMSLE (Working Day)	Train RMSLE (Non Working Day)	Train RMSLE (Average)	Test RMSLE (Working Day)	Test RMSLE (Non Working Day)	Test RMSLE (Average)	Validation RMSLE (Working Day)	Validation RMSLE (Non Working Day)	Validation RMSLE (Average)	Hyperparameters-Working	Hyperparameters-Non Working	Training+Test Time (sec)
ear Regression	0.418155	0.432817	0.422797	0.390881	0.495176	0.428665	0.430658	0.474808	0.444944			0.197
lge Regression	0.423189	0.448141	0.431152	0.394230	0.516847	0.439148	0.438216	0.492361	0.45585	{'alpha': 10}	{'alpha': 10}	1.3
sso Regression	1.313353	1.065252	1.241062	1.285346	1.117959	1.231794	1.31459	1.075364	1.244685	{'alpha': 0.1}	{'alpha': 0.5}	1.3
andom Forest-gorical+Single	0.446963	0.484502	0.459040	0.475814	0.575354	0.511331	0.515769	0.5881	0.539446	n_estimators: 500, max_features: sqrt, min_sam...	n_estimators: 500, max_features: sqrt, min_sam...	5.48
andom Forest-eHotEncoding	0.181866	0.379082	0.260177	0.408674	0.543226	0.458183	0.437411	0.534636	0.470001	n_estimators: 2000, max_features: sqrt, min_sa...	n_estimators: 200, max_features: log2, min_sam...	22.4
andom Forest-Categorical Features	0.322346	0.346449	0.330078	0.409891	0.524065	0.451377	0.43777	0.496375	0.456917	n_estimators: 500, max_features: log2, min_sam...	n_estimators: 500, max_features: sqrt, min_sam...	6.54
Gradient Boosting-eHotEncoding	0.272817	0.299252	0.281356	0.407700	0.549987	0.460304	0.425201	0.558751	0.471078	n_estimators: 3000, learning_rate: 0.01, max_f...	n_estimators: 3000, learning_rate: 0.005, max_...	42.9
Gradient Boosting-Categorical Features	0.312417	0.334973	0.319646	0.387938	0.493577	0.426262	0.404068	0.50143	0.436872	n_estimators: 3000, learning_rate: 0.01, max_f...	n_estimators: 3000, learning_rate: 0.005, max_...	10.1
AdaBoost-eHotEncoding	0.604791	0.558883	0.590809	0.579941	0.651514	0.604868	0.621625	0.599046	0.614649	n_estimators: 5000, learning_rate: 0.001	n_estimators: 5000, learning_rate: 0.001	59.2
Stacking-LR	0.439435	0.510532	0.462858	0.411445	0.491914	0.440051						584.345
Stacking-RF	0.424742	0.475785	0.441349	0.389710	0.498279	0.429161				n_estimators: 1000, max_features: sqrt, min_sa...	n_estimators: 1000, max_features: sqrt, min_sa...	593.277
Stacking-GB	0.411421	0.463945	0.428549	0.395039	0.495763	0.431418				n_estimators: 3000, learning_rate: 0.01, max_f...	n_estimators: 3000, learning_rate: 0.01, max_f...	592.267

Fig 7: Demand Forecasting

10.48047/jocaaa.2025.34.06.18

The RMSLE analysis shows that Linear and Ridge Regression have higher errors, indicating limited ability to capture complex patterns. Lasso performs the worst due to over-penalization. Random Forest and Gradient Boosting achieve better accuracy with lower RMSLE. Stacking models (Stacking-LR, Stacking-RF, Stacking-GB) deliver the best performance, demonstrating that combining models enhances prediction accuracy.

8.Conclusion & Future Scope

In conclusion, the City bike-share usage forecasting tool using historical rental data and predictive analytics represents a significant advancement in urban transportation management. Through rigorous testing and development, the project has demonstrated its capability to accurately predict bike-share demand patterns, optimize resource allocation, and support informed decision-making for city planners, bike-share operators, and users alike.

Future Scope: The future scope of the city bike-share forecasting tool includes enhancing prediction accuracy through advanced models like LSTMs, and GRUs, along with fine-tuned hyper parameters. Incorporating real-time data such as weather, traffic, public events, and economic indicators can make forecasts more adaptive and context-aware. Furthermore, integrating bike-share predictions with public transport systems can promote smarter urban mobility planning and improve overall commuter experience.

References

1. Abouelela, M.; Lyu, C.; Antoniou, C. Exploring the potentials of open-source big data and machine learning in shared mobility fleet utilization prediction. *Data Sci. Transp.* 2023, *5*, 5.
2. Li,X.;Xu,Y.;Zhang,X.;Shi,W.;Yue,Y.;Li,Q. Improving short-term bike sharing demand forecast through an irregular convolutional neural network.*Transp.Res.PartCEmerg.Technol.*2023,*147*,103984.
3. Wang,J. ;Miwa,T.;Morikawa, T. A demand truncation and migration poisson model for real demand inference in free-floating bike-sharing system. *IEEETrans. Intell. Transp. Syst.* 2023,*24*,10525–10536.
4. Lei,Y.;Zhang,J.;Ren,Z.A study on bicycle-sharing dispatching stations it selection and planning based on multivariate data.
5. Zhang, J.; Meng, M.; Wong, Y.D.; Ieromonachou, P.; Wang, D.Z. A data-driven dynamic repositioning model in bicycle-sharing systems. *Int. J.Prod. Econ.* 2021,*231*,107909.
6. Younes,H.;Zou,Z.;Wu,J.;Baiocchi,G. Comparing the temporal determinants of dock

10.48047/jocaaa.2025.34.06.18

less scooter-share and station-based bike-share in Washington, DC.*Transp.Res.PartAPolicyPract.*2020,*134*,308–320.

7. Cantelmo,G.;Kucharski,R.;Antoniou. Low-dimensional model for bike-sharing demand forecasting that explicitly accounts for weather data. *Transp.Res.Rec.*2020,*2674*,132–144 .
8. Gammelli,D.;Peled,I.;Rodrigues,F.;Pacino,D.;Kurtaran,H.A.;Pereira,F.C. Estimate latent demand of shared mobility through censored Gaussian processes.*arXiv*2020,arXiv:2001.07402.
9. Durán-Rodas, D.; Chaniotakis, E.; Wulfhorst, G.; Antoniou, C. Open source data-driven method to identify most influencing spatiotemporal factors. An example of station-based bike sharing. In *Mapping the Travel Behavior Genome*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 503–526.
10. Yang,Y.; Heppenstall, A.;Turner, A.;Comber,A. Using graph structural information about flows to enhance short-term demand predictioninbike-sharingsystems.*Comput.Environ.UrbanSyst.*2020,*83*,101521.
11. Sathishkumar,V.E.;Cho,Y.Season wise bike sharing demand analysis using random forest algorithm.*Comput.Intell.*2020*40*,e12287.
12. Boufidis, N.; Nikiforiadis, A.; Chrysostomou, K.; Aifadopoulou, G. Development of a station-level demand prediction and visualization tool to support bike-sharing systems' operators. *Transp.Res.Procedia*2020,*47*,51–58.
13. Jin,H.;Jin,F.;Wang,J.;Sun,W.;Dong,L.Competition and cooperation between shared bicycles and public transit: A case study of Beijing. *Sustainability*2019, *11*, 1323.
14. Zheng, F.; Gu, F.; Zhang, W.; Guo, J. Is bicycle sharing an environmental practice?Evidence from a life cycle assessment based onbehavioral surveys. *Sustainability* 2019, *11*, 1550.
15. Okraszewska, R.; Romanowska, A.; Wołek, M.; Oskarbski, J.; Birr, K.; Jamroz, K. Integration of a multilevel transport system model into sustainable urban mobility planning. *Sustainability*2018, *10*, 479.
16. Lei,C.;Ouyang,Y. Continuous approximation for demand balancing in solving large-scale one-commodity pickup and delivery problems. *Transp. Res. Part B Methodol.* 2018, *109*, 90–109.
17. Xu,C.;Ji,J.;Liu,P. The station free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transp. Res. Part CEmerg.Technol.*2018,*95*,47–60.

10.48047/jocaaa.2025.34.06.18

18. Lin, L.; He, Z.; Peeta, S. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transp.Res.PartCEmerg.Technol.*2018,*97*,258–276.
19. Lin, P.; Weng, J.; Liang, Q.; Alivanistos, D.; Ma, S. Impact of weather conditions and built environment on public bike sharing trips in Beijing. *Netw. Spat. Econ.* **2018**, *20*, 1–17.
20. Shen,Y.;Zhang,X.;Zhao,J. Understanding the usage of dock less bike sharing inSingapore.*Int.J.Sustain.Transp.***2018**,*12*,.686–700