DOI: 10.15439/2023R87
ISSN 2300-5963 ACSIS, Vol. 38

AI-driven rental bicycle system: An Ensemble learning approach

Cu Kim Long
Information Technology Center Ministry of Science and Technology (MOST)
Hanoi, Vietnam longck@most.gov.vn

Trinh Thi Thu Hoang
Hanoi University of Industry
Hanoi, Vietnam huongttt@haui.edu.vn

Gloria Jeanette Rinco´n Aponte
Universidad Cooperativa de Colombia
Bogota, Colombia
gloriaj.rincon@campusucc.edu.co

Vikram Puri
Center of Visualization and Simulation
Duy Tan University
Da Nang, Vietnam
purivikram@duytan.edu.vn

***Abstract*—Bicycle sharing is a notable sustainable transporta- tion option for metropolitan regions and communities seeking to address environmental concerns, reduce traffic congestion, and combat air pollution while promoting public health and improving connections. There are already technologies to support this system, including typical mobile applications and kiosks strategically positioned at the bicycle station. Nevertheless, most proposed solutions cannot accurately forecast the demand for bicycle availability, efficiently redistribute bicycles, create routes to circumvent traffic congestion and conduct comprehensive user analysis. To address these challenges, a framework for an AI- enabled bicycle-sharing system has been presented to predict the count of bicycle rentals. To assess performance, four distinct ensemble-based models are implemented and tested using various statistical parameters.**

***Index Terms*—bicycle rental system, BSS, artificial intelligence, ensemble technique, feedback.**

I. Introduction

Now, there exists an escalating worldwide inclination towards the adoption and execution of bicycle rental systems [1]. The main goal of the bicycle rental system is to enable the tem- porary leasing of bicycles to individuals, typically for periods spanning from 15 minutes to a few hours. There are some key reasons to escalate the demand for the bike-sharing system, such as providing a sustainable transport option which en- courages people to use bicycles instead of fuel-based vehicles, providing flexibility for pick-up and dropping bikes due to nu- merous docking stations, reducing traffic congestion inside the cities, connected with the public transportation facilities which make it convenient to pick up public transportation, offer a convenient way for the tourists to explore the city more as well as promote tourism activities. Additionally, it also helps improve people's physical and mental health [2]. Although the bike rental system greatly impacts society, key issues also need to be addressed. The first issue is bike availability; at peak times, bikes are unavailable at their docking station, discouraging people from adopting the bike-sharing system. Large-scale bike upkeep and repairs require a lot of work and skill. Systems must install more stations and determine the optimal fleet composition to handle rising demand over time [3]. Understanding user patterns and peak usage periods is crucial to adapting to changing needs despite difficulty. Pricing methods must balance income generation with affordability to maintain the system. Few trip planning tools point users toward safe, efficient cycling routes. Predicting maintenance issues before malfunctions render bikes useless is one unsolved topic. Keeping an eye on fleet activities, bike maintenance, customer behaviour, pricing, routes, and breakdowns is challenging [4]. Artificial intelligence (AI) has the potential to overcome these challenges and provide a better experience for users as well as service providers. AI encompasses computer systems that possess the ability to execute tasks that conventionally necessitate human intelligence, including but not limited to sensory seeing, recognition of speech, decision- making, and translating languages [5]. The fusion of AI in the bike-sharing system transforms the user experience better. The key benefits of AI in bike-sharing systems are (see in figure 1):

1) **Demand Prediction:** AI analyses the bike availability per the rider data log in the system, including the external weather conditions, and provides feedback to service providers to rebalance the bike availability.
2) **Optimization in Routes:** As per traffic congestion in the city, the system optimizes new routes and provides the fastest and safest way for the user.
3) **Maintenance:** AI can analyze data collected from bike sensors to detect bicycles that require repair or main- tenance before the occurrence of any breakdowns. This enhancement contributes to the enhancement of safety and reliability.
4) **Rewards:** AI can comprehend user data and offer incen- tives, such as points and monetary prizes, to encourage prolonged user engagement with bike-sharing systems.
5) **Fraud Detection:** AI can identify and analyze usagepatterns that may indicate questionable activity, hence speeding up identifying stolen or lost bicycles. This phenomenon leads to a decrease in both financial losses and criminal activities.

This study presents a potential architecture for a bike rental system incorporating artificial intelligence technology. The primary contribution of this work is as follows:

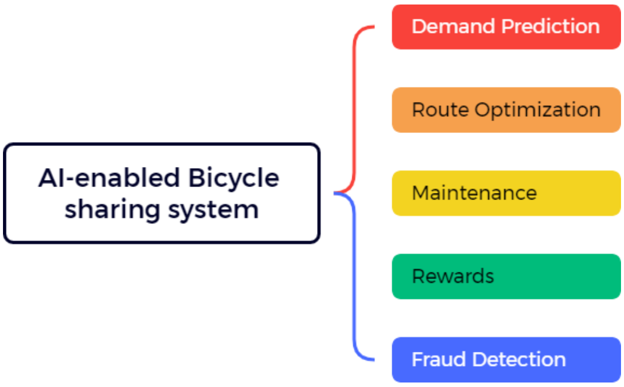

Fig. 1. Key points to integrate the AI with Bicycle sharing system

1) Trained four ensemble-based approaches on BSS set.
2) Evaluated these models with the help of statistical parameters.
3) chose and deployed a better performance model.

The subsequent sections of the paper are organized as follows: Section 2 provides an overview of the associated work, Section 3 outlines the technique employed in the suggested work, Section 4 presents the results obtained from the proposed work, and Section 5 summarizes the study.

II. Related Work

In the past years, some studies have been published on bicycle-sharing systems. In [6], the authors suggested using a multiple-layer spatial network model to analyze the public transport system. This model considers the interconnectivity of transit paths, cycling stations, and pedestrian pathways. The overall evaluation of this study is that the introduction of public bicycle-sharing systems has been shown to improve the efficiency of the public transportation network by decreasing passenger travel times, promoting smoother traffic flow, and alleviating congestion. In this study [7], authors present a complete technique for establishing a Bicycle Sharing System (BSS) that efficiently incorporates optimizing station positions and capacity allocation. The proposed methodology combines a set-covering framework for distributing customer demand to stations with a queuing model for assessing amenities levels. These studies are manually optimizing station location and capacity location. AI can enhance the design, operation and adaptability of the BSS. To determine which machine- learning techniques are most commonly used in this field and to examine how machine learning has been used to enhance bike-sharing programs in smart cities, authors [8] presented a literature review. The review aids in synthesizing prior findings and identifies areas needing additional investigation. In [9], the authors presented how to deploy ML models to maximize the number of bicycles available in the public bicycle-sharing scheme. The algorithms provide an accurate forecast of station occupancy levels, enabling timely redis- tribution of bicycles across stations. Ensuring riders' regular access supports bicycles as an environmentally friendly form of travel for the environment and public health. The deep learning approach presented [10] a highly accurate forecasting model that has the potential to significantly contribute to real-time decision-making and operational management within expanding dockless systems worldwide. The main aim of this study is to expand the application of advanced neural network architectures in the modelling of complex spatiotemporal systems. Similarly, the authors [11] proposed a deep learning model called STGA-LSTM to forecast the demand for bicycle sharing across several stations. Advocating for fair and just consumption practices facilitates the establishment of envi- ronmentally friendly, reliable shared transportation systems. The authors suggest a dynamic repositioning system based on a Monte Carlo tree search [12]. This system aims to assist service providers in efficiently balancing the distribution of bicycles across stations, taking into account their movement patterns. The notion of service level is established to quantify the quantity of bicycles that require transfer at each station. In [13], this study primarily applies deep learning models to forecast short-term bike demand for the bike rental system, specifically predicting demand 15 minutes in advance. A hybrid CNN-LSTM model is considered for the prediction.

III. Methodology

In this section, the methodology of the paper, including the framework and how things are connected to the framework, is discussed (see figure 2).

A. *Bicycle Sharing Ecosystem*

The Bicycle Sharing Ecosystem (BSS) allows individuals to borrow bicycles within their local area temporarily. A considerable number of bicycles are observed to be parked at a specific station inside the BSS. Bicycles can be connected through docks, which are specialized racks designed to secure and release the bike, or they can be supplied with specialist locks that keep the bicycle stationary at a certain location. Furthermore, these locks can be operated by mobile applica- tions or kiosks located at the terminal. The mobile applications examined in the BSS are founded on the conventional server- client infrastructure for reserving and verifying the availability of bicycles. In this system, an AI system is integrated into the traditional mobile application to forecast the total number of bicycles used.

B. *AI Model Technique*

To train and deploy the AI model in the BSS, there are some steps which need to be followed.

1) *Data Collection and Data Pre-processing*

The dataset [14] on bike sharing encompasses many characteristics about the date, time, weather conditions, day type, and the count of bike rentals. The variables encompass the record index, date, season, year, month, hour of the day, presence of a holiday, day of the week, and working day status. The variable "weathersit" classifies weather conditions into four distinct categories: clear, mist, light, and heavy rain. The dataset additionally has normalized variables for temperature, perceived temperature, humidity, and wind speed. In conclusion, the dataset includes tallies about the number of individuals classified as casual riders registered riders and the overall count of bike rentals, encompassing casual and registered users. These variables collectively offer valuable insights into the various aspects that affect the

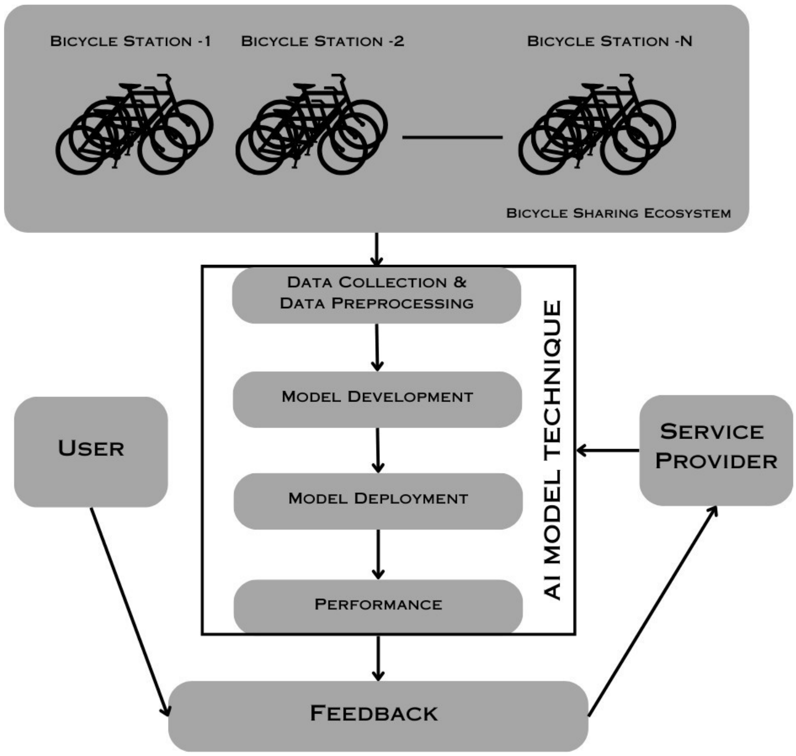

Fig. 2. Framework of AI-enabled Bicycle sharing system

demand for bike rentals, such as weather conditions, time of day, kind of day (e.g., weekday or weekend), and consumer characteristics. The objective is to utilize these factors to investigate and construct a model for the patterns of bike sharing.

Data Pre-processing: The data pre-processing techniques [15] include removing the missing values using mean, median or most frequently values imputation; the next step is to identify the abnormal values that need to be removed from the dataset. Feature selection is a crucial step in the modelling process, as it involves identifying and removing duplicate, unnecessary, or noisy features. By selecting those with the greatest importance characteristics, the precision and generality of the model can be enhanced.

2) Model Development

To facilitate the creation of the model, it is necessary to partition the data into either a training set and a testing set or into a training set, a testing set, and an evaluation set. The rationale behind withholding specific test data is to ensure an impartial assessment of the model's ability to generalize novel variables. The outcome evaluation may demonstrate a positive bias if the model is assessed using the same dataset used for its training. The subsequent phase involves selecting an appropriate model following the dataset. One approach that can be employed is to utilize various models and assess their performance by employing statistical parameters. This methodology facilitates the provision of feedback to the model under consideration. In this study, four ensemble machine learning are deployed to evaluate the performance of the proposed approac and explained in Table I.

3) Model Deployment

Once the model has been trained, it is integrated into the manufacturing ecosystem, where it may generate predictions in real-time. It is imperative that the input information format and sort utilized for production purposes align with the format and type on which the algorithm was developed [16]. The set-up system must be capable of effec- tively managing the pre-processing of real-time data. For this study, the proposed model is deployed through FastAPI.

TABLE I. DIFFERENT ENSEMBLE MODELS USED IN THE STUDY

S.No.	Model Ensemble	Description
1	Random Forest (RF) Regressor	The RF algorithm is a metaestimator that employs several decision trees for classification. RF model is a versatile, easy to use regression model that provides better accuracy without extensive hyperparameter tuning.
2	Gradient Boosting (GB) Regressor	GB estimator constructs an additive model using a forward stage-wise approach, enabling the optimal selection of various distinct losses. Although GB model provides accurate prediction, it requires careful tuning.
3	AdaBoost Regressor	The AdaBoost is a meta-estimator that initially trains a regression model on the initial data set. Sub- sequently, it trains new regressor clones on the exact same dataset, but with updated instance values based on the variance of the present prediction. Additionally it required the tuning to avoid the overfitting issue.
4	Extra Tree Regressor	This model presents a meta estimator that applies a series of randomized decision tree structures, also known as extra-trees, on different subsets of the dataset

4) Performance

The performance of AI-based applications is of utmost importance due to the tendency of models to deteriorate in intricate real-world settings as time progresses.

In order to mitigate the impact of errors on consumers, it is imperative to engage in ongoing surveillance of essential per- formance indicators, thereby enabling the timely identification of potential difficulties before they escalated into significant problems. This enables the model to undergo retraining or augmentation over time. Some parameters are also helpful to check the model performance after deployment, such as prediction accuracy, error rate, AI pipeline monitoring, and Quinton sampling, which helps to check the model perfor- mance through the experts in a periodic period.

C. Feedback

The incorporation of feedback is widely recognized as an essential component of the machine-learning process. The feedback provided by real-time users holds significant impor- tance as it is a crucial indicator of the model's performance. This is because even if a model demonstrates a high validation score, it may still encounter failures when deployed in real- world scenarios. In certain instances, it may be the case that researchers and developers are unable to discern the fault within the code. However, the user can pay attention to and discover said error. Furthermore, feedback mechanisms foster trust and promote transparency on the capabilities and limitations of models [17]. In this research, participants and service providers are linked to the feedback mechanism of the artificial intelligence model. The user submits a problem using the feedback system, which is relayed to the service providers for prompt resolution

IV. Results and Evaluation

In this section, experimental testbed and results outcome is discussed.

A. Testbed

For this study, google co-lab is considered to train the model initialized with matplotlib, pandas, NumPy, seaborn and sklearn for the training and testing of four ensemble model for this study.

B. Analysis

To evaluate the performance evaluation of the proposed framework, there are three statistical parameters [18] considered such as root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R2).

Table II displays the performance evaluation of four ensemble models on the BSS dataset. A dataset was utilized to assess the performance of four regression ensemble models in predicting a continuous target variable. The models included in the training process included the Random Forest Regressor, Gradient Boosting Regressor, AdaBoost Regressor, and Extra Trees Regressor. Smaller MAE and RMSE values are indica- tive of superior model performance. Higher R2 values that approach 1 are indicative of superior performance. The Extra Trees Regressor demonstrated superior performance with a mean absolute error MAE of 24.87, RMSE of 41.17, and R2 of 0.94. This demonstrates that it possesses the lowest mean errors and exhibits the highest degree of concordance with the actual target values. The Adaboost Regressor had the lowest performance across all three evaluation measures, namely a MAE of 86.13, RMSE of 106.62, and R2 value of 0.64. The predictions exhibit the highest degree of deviation from the actual data. The RF and GB variants demonstrated a moderate level of performance. The errors and R2 values of their models fall within the range of the best and worst models. Based on the analysis conducted, it is recommended to utilize the Extra Tree Regressor model for predicting the target variable due to its notable performance across many assessment measures. The measurements offer empirical support for the selection of this option, ensuring its accuracy.

Table II. Performance Evaluation of Four Ensemble Model on BSS Set

Parame-ter	**RF Regressor**	**GB Refressor**	**AdaBoost Regressor**	**Extra Tree Regressor**
MAE	25.37	48.18	86.13	24.87
RMSE	42.05	70.78	106.62	41.17
R^2	0.94	0.84	0.64	0.94

The objective of regression modeling is to make predictions for a continuous target variable by utilizing a collection of pre- dictor variables. The regression model produces forecasts, or estimated quantities, of the dependent variable. The aforemen- tioned predictions are indicative of the results produced by the regression model that has been appropriately fitted, based on a certain set of input predictor values. The comparison between the predicted counts derived from the regression model and the actual counts is conducted to assess the effectiveness and accuracy of the fitted model. The accuracy of the model in predicting the target variable improves as the projected counts approach the actual counts. An effective regression model aims to minimize discrepancies and achieve predicted numbers that closely align with the actual counts. Figure 3 illustrates the comparison between the observed and anticipated total count of bicycles for four different ensemble learning models. Figure 3. actual vs predicted Total Count figure with two columns. Left column: (A) Top left panel showing RF Regressor. (B) Bottom left panel showing AdaBoost Regressor. Right column: (C) Top right panel showing GB Regressor. (D) Bottom right panel showing Extra Tree Regressor.

V. Conclusion

Bicycle-sharing systems offer an ecologically advantageous mode of transportation that confers numerous benefits to metropolitan regions regarding environmental sustainability. Nevertheless, the efficient operation of these systems poses significant logistical and planning challenges. The increasing significance of AI technologies is observed in the context of bicycle-sharing providers, who are utilizing these tools to tackle the challenges mentioned earlier. AI algorithms can examine patterns in rider utilization to optimize bike allocation and determine appropriate pricing strategies. This paper proposes an AI-enabled framework for analysing rental bicycles in a bicycle rental system. The framework aims to enhance the service provider's understanding of rental bicycles and improve the customer experience.

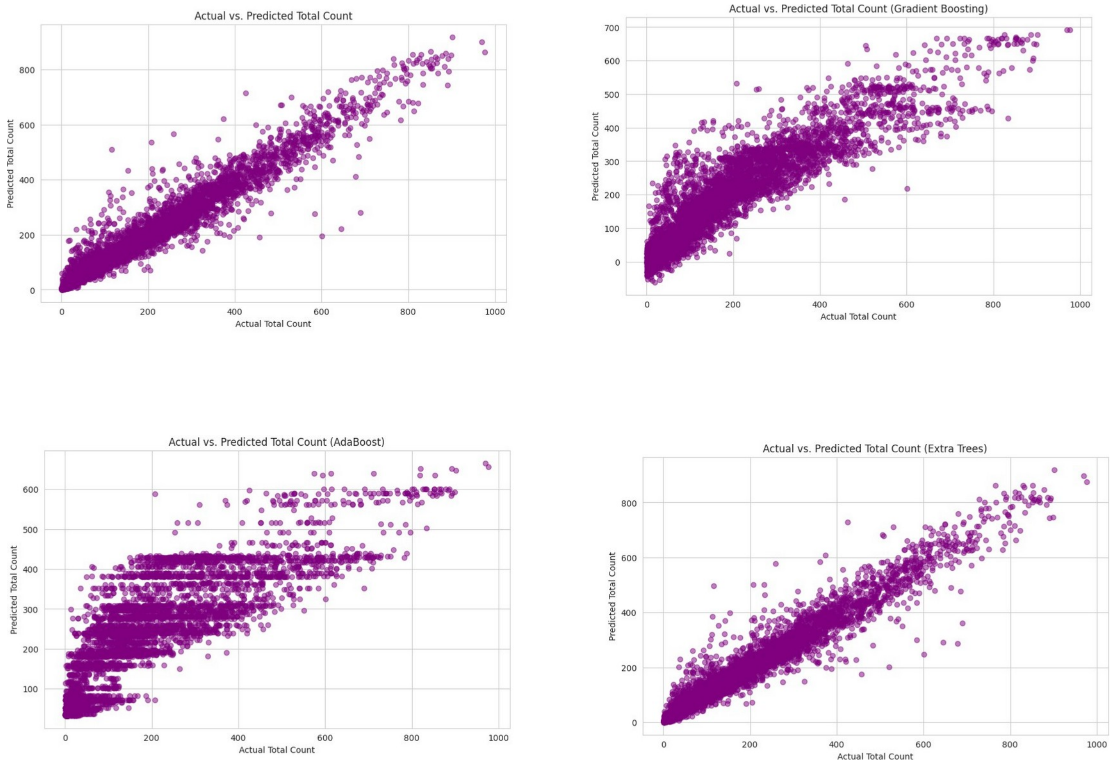

Fig. 3. Actual vs Predicted Total Count bicycle for four ensemble learning approaches.

There is a significant potential for enhancing intelligence and efficiency within bicycle-sharing systems in the foreseeable future by integrating AI technology to a greater extent. The use of advanced tracking sensors and predictive analytics will augment the accuracy of monitoring the whereabouts of bicycles and forecasting ridership demand. AI methodologies enable service providers to offer individualized recommendations and incentives to specific clients. In future times, this technology is expected to expand its range to include intelligent route planning, while simultaneously enhancing the user experience.

References

[1] Billhardt, H., Ferna´ndez, A., & Ossowski, S. (2021). Smart recommen- dations for renting bikes in bike-sharing systems. Applied Sciences, 11(20), 9654.

[2] Czech, P., Turon´, K., & Urban´czyk, R. (2018). Bike-sharing as an element of integrated Urban transport system. In Advanced Solutions of Transport Systems for Growing Mobility: 14th Scientific and Technical Conference” Transport Systems. Theory & Practice 2017” Selected Papers (pp. 103-111). Springer International Publishing.

[3] Ma, Y., Lan, J., Thornton, T., Mangalagiu, D., & Zhu, D. (2018). Challenges of collaborative governance in the sharing economy: The case of free-floating bike sharing in Shanghai. Journal of cleaner production, 197, 356-365.

[4] Shaaban, K. (2020). Why don’t people ride bicycles in high-income developing countries, and can bike-sharing be the solution? The case of Qatar. Sustainability, 12(4), 1693.

[5] Duan, Y., & Wu, J. (2022). An AI Approach to Rebalance Bike- Sharing Systems with Adaptive User Incentive. Artificial Intelligence-based Internet of Things Systems, 365-389.

[6] Yang, X. H., Cheng, Z., Chen, G., Wang, L., Ruan, Z. Y., & Zheng, Y. J. (2018). The impact of a public bicycle-sharing system on urban public transport networks. Transportation research part A: policy and practice, 107, 246-256.

[7] C¸elebi, D., Yo¨ru¨su¨n, A., & Is¸ık, H. (2018). Bicycle sharing system design with capacity allocations. Transportation research part B: method- ological, 114, 86-98.

[8] Albuquerque, V., Sales Dias, M., & Bacao, F. (2021). Machine learning approaches to bike-sharing systems: A systematic literature review. ISPRS International Journal of Geo-Information, 10(2), 62.

[9] Rozˇanec, J. M., Krivec, T., Kersˇicˇ, V., Cundricˇ, L., Stojanovicˇ, B., Zeman, M., & Bratko, I. (2022). Bicycle Sharing Systems meet AI: forecasting bicycles availability and decision-making. In Central European Conference on Information and Intelligent Systems (pp. 365-370). Faculty of Organization and Informatics Varazdin.

[10] Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D., Chen, W., & Ju, Y. (2019). A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. Neural Computing and Applications, 31, 1665-1677.

[11] Ma, X., Yin, Y., Jin, Y., He, M., & Zhu, M. (2022). Short-term prediction of bike-sharing demand using multi-source data: a spatial-temporal graph attentional LSTM approach. Applied Sciences, 12(3), 1161.

[12] Huang, J., Tan, Q., Li, H., Li, A.,& Huang, L. (2022). Monte carlo tree search for dynamic bike repositioning in bike-sharing systems. Applied Intelligence, 1-16.

[13] Mehdizadeh Dastjerdi, A., & Morency, C. (2022). Bike-sharing demand prediction at community level under COVID-19 using deep learning. Sensors, 22(3), 1060.

[14] Fanaee-T,Hadi. (2013). Bike Sharing Dataset. UCI Machine Learning Repository. https://doi.org/10.24432/C5W894.

[15] Zhang, C., Wu, F., Wang, H., Tang, B., Fan, W., & Liu, Y. (2022). A meta-learning algorithm for rebalancing the bike-sharing system in IoT smart city. IEEE Internet of Things Journal, 9(21), 21073-21085.

[16] Ehsan, A., Abuhaliqa, M. A. M., Catal, C., & Mishra, D. (2022). RESTful API testing methodologies: Rationale, challenges, and solution directions. Applied Sciences, 12(9), 4369.

[17] Li, J., & Wang, W. (2023). From Renting Economy to Sharing Economy: How Do Bike-Sharing Platforms Grow in the Digital Era?. Journal of the Knowledge Economy, 1-21.

[18] Kataria, A., & Puri, V. (2022). AI-and IoT-based hybrid model for air quality prediction in a smart city with network assistance. IET Networks, 11(6), 221-233