

## Research Article

# Electric Kickboard Demand Prediction in Spatiotemporal Dimension Using Clustering-Aided Bagging Regressor

Prince Waqas Khan , Se-Joon Park , Sang-Joon Lee , and Yung-Cheol Byun 

*Department of Computer Engineering, Jeju National University, Jeju-si, Republic of Korea*

Correspondence should be addressed to Sang-Joon Lee; [sjlee@jejunu.ac.kr](mailto:sjlee@jejunu.ac.kr) and Yung-Cheol Byun; [ycb@jejunu.ac.kr](mailto:ycb@jejunu.ac.kr)

Received 5 February 2022; Revised 29 March 2022; Accepted 20 July 2022; Published 30 August 2022

Academic Editor: Mohammad Miralinaghi

Copyright © 2022 Prince Waqas Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Demand for electric kickboards is increasing specifically in tourist-centric regions worldwide. In order to gain a competitive edge and to provide quality service to customers, it is essential to properly deploy rental electric kickboards (e-kickboards) at the time and place customers want. However, it is necessary to study how to divide the region to predict electric mobility demand by region. Therefore, this study is made to more accurately predict future demand based on past regional customers' electric mobility demand data. We have proposed a novel electric kickboard demand prediction in spatiotemporal dimension using clustering-aided bagging regressor. We have used electric kickboard usage data from a Jeju, South Korea-based company. As a result of the experiment, it was found that the accuracy before using clustering-based bagging regressor and when the region was divided by the clustering method, the performance was improved, and we have achieved a regression score ( $R^2$ ) of 93.42 using our proposed approach. We have compared our proposed approach with other state-of-the-art models, and we have also compared our model with different other combinations of bagging regressors. This study can be helpful for companies to meet the user's demand for a better quality of service.

## 1. Introduction

With the increasing demand for fuel-efficient vehicles, due to growing concerns about greenhouse gas and carbon emissions, the use of electric kickboards and scooters is expected to increase over time. Since 2016, the penetration index of this sharing-service acceptance rate has been growing [1]. In addition, the need for sustainable urban mobility and modern transportation infrastructure is leading to a shift from traditional modes of transport to electronic modes of transport. Demand for this electric mobility has grown significantly. An electric kickboard is a two-vehicle device. People can ride it while standing on it. It is suitable for one and a maximum of two persons. Figure 1 shows a picture of an electric kickboard.

Electric kickboards (e-kickboards) are expected to affect energy security and air quality positively. As industry 4.0 develops, companies have begun to study how to use big data, which is a collection of previous data, to meet customer

needs. Such big data only applies to companies that have continued to operate, and for startups starting a new business, most of the data is small in amount and unstable in the form of income [3]. It is difficult to meet customer needs with insufficient qualitative and quantitative data. Customers who want to use electric mobility feel satisfied with using the service without being restricted by time and place when they want to use it. However, suppose it cannot be used due to finite electric mobility. In that case, customers may feel inconvenient, and if this case continues for companies, demand will decrease, which may affect sales [4]. That is why companies need to place them where and when customers want them.

This study used data from an e-kickboards company that provides electric mobility services in Jeju Island, South Korea. Figure 2 shows the location of kickboard stations in Jeju Island, South Korea. Tourists or local residents can rent a kickboard from these stations. Since there is nothing more important than safety, the company also provides free



FIGURE 1: Electric kickboard [2].

helmets and safety gear. The top speed of the kickboard is 25 km/h.

Deploying through demand forecasting may introduce errors. If electric mobility is insufficiently deployed, this gap can be filled by bringing it from where there is a surplus of demand. However, the longer the distance between the two regions, the longer it will take to serve customers. Based on this idea, this study can also enhance prediction results by dividing the regions in Jeju Island, placing electric mobility with remaining demand in the center of each region, and quickly relocating where errors occur. In this article, when dividing the area in Jeju Island into several places, nearby rental stations were bundled based on the centroid, and electric mobility demand prediction was carried out. Electric mobility is stored in the center of each region. If a region with more demand than predicted through regional forecasting occurs, quick service can be provided by fetching and deploying electric mobility in the region's center. In addition, it was confirmed that the prediction accuracy was

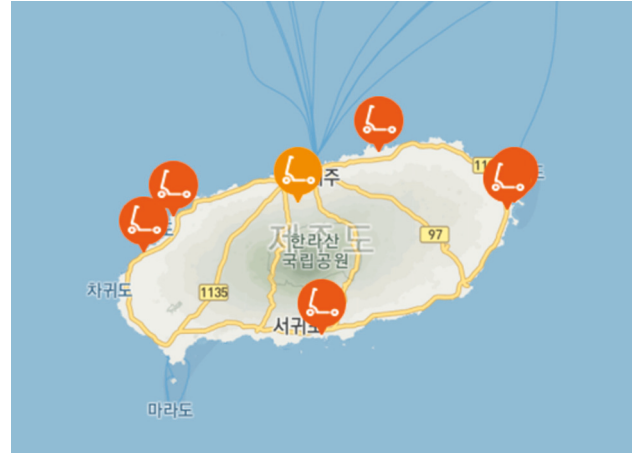


FIGURE 2: Rental locations of kickboard in Jeju Island.

increased when the learning results were compared before and after dividing the regions.

**1.1. Contributions.** We proposed to use a clustering-aided bagging regressor for electric kickboard demand prediction in spatiotemporal dimensions. We used the  $k$ -means algorithm to cluster e-kickboard demand areas and identify the centroid and then used a bagging ensemble for demand prediction. The proposed bagging ensemble consists of a base layer and metalayer. The base layer consists of XGBoost, Extra Trees regressor, and Random Forest model, whereas the metalayer consists of an Extra Trees regressor. The significant contribution of this article can be summarized as follows:

- (i) Integrating spatial, temporal, and weather data to cover the effect of different parameters on the mobility demand
- (ii) Employing data and predictive analytics techniques for data obtained from electric kickboard company
- (iii) Integrating clustering and bagging regressor to develop a hybrid prediction model to predict e-kickboard demand
- (iv) Comparing the proposed model with state-of-the-art ML models and different combinations of ensemble models

The rest of the article is arranged as follows. Section 2 presents the relevant approaches and publications and compares existing literature and proposed methodology. Section 3 introduces the methodology used in this research study. Section 4 presents an analytical and graphical analysis of data. Section 5 provides results and covers comparisons with the latest machine learning models. Finally, we conclude the article in Section 6.

## 2. Related Work

Different machine learning models have been used by researchers, such as long short-term memory (LSTM), Generalized Autoregressive Conditional Heteroskedasticity

(GARCH), Seasonal Autoregressive Moving Average (SARIMA), Bidirectional Encoder Representations from Transformers-Based Deep Spatial-Temporal Network (BDSTN), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), hexagonal convolution operation LSTM (H-ConvLSTM), and Dynamic Spatiotemporal GNN method with Tensor Network (DSTGNN). Table 1 shows the contributions of this research in the context of the literature.

**2.1. E-Mobility Data Analysis Using Deep Learning.** Greenhouse gas emissions (GHG) and high fuel consumption have become a significant issue these days [14]. In particular, CO<sub>2</sub> emissions from transportation have reached almost a quarter of global emissions [15]. The electric vehicle (EV) is considered an exciting alternative to solve the above problems [16]. However, issues such as inadequate billing infrastructure have also arisen to support the growing demand in the growing EV market. Effective forecasting of commercial EV bill demand ensures the reliability and stability of short-term network utilities and supports investment planning and resource allocation for long-term infrastructure bills. The article by Yi et al. [17] provides an overview of the monthly commercial EV load application time series using the deep learning approach. The proposed model was confirmed by original datasets in Utah and Los Angeles. Two forecasting purposes, one-step forward forecasting and multistep forward forecasting, were reviewed. In addition, the model was compared to other time series and machine learning models. Experiments show that both Seq2seq and LSTM provide satisfactory one-step predictions. However, when making multistep predictions, the Seq2seq outperforms other models in terms of various performance measurements, demonstrating the model's powerful ability to predict sequential data.

The study by Zuniga-Garcia et al. [18] considers the issue of the use of travel data in the mobility data specification standard for meaningful analysis of the use of the infrastructure of e-scooters without the use of personally identifiable numbers. They examine the integration of e-scooters into the city infrastructure of Austin, Texas, using e-scooter supplier travel data and infrastructure geographic inventory data. Their analysis shows that e-scooters use about 80,000 e-scooter rides each year, more than 11 million, which is 1.4 percent of the total electric scooter rides in the city during this period. Their results show that the average distance traveled by electric scooter was divided between pedestrians (18 percent), bicycle lanes (11 percent), roadside (33 percent), and other nonspecific locations (38 percent). In addition, about sixty percent of road trips were made on large arteries, and bicycle lane users prefer moderate to high comfort levels. The purpose of the article by Davies et al. [19] is to examine the current distribution and development challenges of tourist destinations, focusing on micromobility. Micromobility is linked to a new model that includes but is not limited to hiking, cycling (both current modes), e-bikes, and e-scooters (new modes). The proliferation of new micromobility in destination urban areas can be

viewed positively in terms of their sustainable urban dynamics, thus increasing the chances of attracting tourists. However, it is also pessimistic about potential issues with space, accessibility, and sustainable implementation. Therefore, destination developers and partners need to consider successfully integrating micronavigation into a sustainable transportation system.

**2.2. E-Scooter Demand Prediction.** The study by Bai and Jiao [20] provides empirical evidence of e-scooter travel in two US cities. Moreover, transportation and training literature compares the two cities and emphasizes the importance of local individuality. The results of the regional analysis show that electric scooters were widely used in the city center and on the university campus. However, patterns of electric scooter use in both cities were temporarily different: Austin faces more electric scooter congestion during the day and weekends, while Minneapolis shows large rides at night and throughout the week. In the article by Feng et al. [21], they used a large amount of different Twitter data—including text; references; GPS data; general images; and e-scooter app screenshots, emojis, and emoticons—to analyze electric scooter racing services. Over the past 18 months, more than 5 million English tweets referring to the word “scooter” or scooter emoji have been added. They first did extensive data preprocessing to eliminate noise and reduce false positives. They believe that the results obtained by the public will provide a deeper insight into the emergence of e-scooter services as a generic directory in smart cities. The study by Kolaković-Bojović et al. [22] provides a quantitative and qualitative analysis of the data by pointing out the critical issues presented in both newspapers and Twitter posts. They conducted media readings on its effects on various aspects of its environmental well-being. The authors examine, among other things, the use of electric scooters in the press and the challenges to civic security, as well as the relationship between the posts of the Twitter community in Serbia. In general, they tried to answer whether electric scooters can be considered a security challenge in the city or any other issue of moral fear.

**2.2.1. Bus Arrival and Departure Time Prediction.** There is always some uncertainty about public buses' arrival and departure times, such as signals, bus stop times, climate change, and fluctuations in travel requirements [23]. In developing countries, this uncertainty is heightened by the availability of redundant vehicles, different modes of transportation, and a lack of risk discipline. Therefore, the issue of forecasting remains a challenge, especially with the arrival of buses in developing countries. The work by Achar et al. [24] suggests a new way of predicting the arrival of buses in real time. Unlike the current method, the proposed method learns transport interactions and patterns. It first recognizes the unknown sequence of spatial dependencies and then understands the linear, non-static spatial correlation for this discovered sequence. It retains the temporal relationship between continuous travel and changing time. The learned prediction model is rewritten in an appropriate

TABLE 1: Comparison of recent research contributions for vehicle demand prediction.

Sr#	Ref, year	Clustering	Regression	Proposed model	Vehicle type	Weather	Area
1	[5], 2018	Yes	No	Hierarchical clustering	E-scooter	No	Germany
2	[6], 2019	No	Yes	LSTM	Taxi	Yes	Xiamen and Chengdu, China
3	[7], 2020	No	Yes	GARCH, SARIMA	E-scooter	No	Thammasat University, Thailand
4	[8], 2021	No	Yes	BDSTN	Taxi	Yes	New York, USA
5	[9], 2021	Yes	No	HDBSCAN	E-scooter	No	Berlin, Germany
6	[10], 2022	Yes	Yes	LSTM	E-scooter	Yes	Seocho and Gangnam, South Korea
7	[11], 2022	No	Yes	Random Forest	Car, bicycle	No	Hamburg and Hanover, Germany
8	[12], 2022	No	Yes	H-ConvLSTM	Car	No	Chengdu, China
9	[13], 2022	No	Yes	DSTGNN	Bike, taxi	No	New York, USA
10	Proposed	Yes	Yes	Bagging regressor	E-kickboard	Yes	Jeju, South Korea

linear state-specific format to make the best predictions come true, and then the Kalman filter is applied. Performance was analyzed using actual field data and compared with existing methods.

**2.2.2. EV Power Demand Prediction.** The study by Tianheng et al. [25] outlines the power demand forecasts for electric vehicles and the strategy for overcoming errors in forecasting. The goal is to reduce fuel consumption in real-time operations. They develop a neural network model to estimate the vehicle's electricity demand. Furthermore, a mathematical model is proposed to convert the predicted power demand into a battery charge status reference, greatly simplifying the charging programming system. Finally, they use the adaptive equivalent consumption minimization strategy to monitor referrals and determine the status of the propulsion system. The proposed approach enables the maximum distribution of power between engines and cars worldwide and maximum torque distribution locally. The simulation power-sharing plug-in was performed on a hybrid electric bus. Compared to rule-based and proposed technologies, the proposed method has significantly improved fuel consumption and other indicators.

**2.3. Taxi Demand Prediction.** Taxi demand forecasts have recently attracted increased research interest. In the article by Liu et al. [26], they presented challenging and award-winning work entitled Taxi Origin-Destination Demand Forecast, which aims to predict the demand for taxis across all regions at a later interval. The critical challenge was effectively acquiring various relevant kinds of information to learn the types of questions. They addressed this issue using the new contextual spatial time network to model the local spatial context, the temporary evolutionary perspective, and the global communication perspective. Extensive testing and analysis of a large database demonstrated the high efficiency of their contextual spatial time network model compared to other methods for predicting the actual destination needs. Real-time and precise taxi demand forecasts can help drivers in booking taxi resources for the city in advance, help drivers find passengers faster, and reduce waiting times. Many current studies have focused on the local and temporal characteristics of taxi demand distribution, with a lack of a model of the relationship between taxi pick-up demand and

download demand from a multipurpose learning perspective. In the article by Zhang et al. [27], they proposed a multifunctional learning model with three parallel LSTM levels for predicting and downloading taxi demand and the single demand for performance prediction methods and two on-demand prediction methods. Experimental results from datasets show that the demand for collection and the demand for downloads depend on each other and the accuracy of the suggested co-deduction system.

Taxi demand forecasting plays an important role, especially in ranking resources to help differentiate between demand and service in times of economic sharing and autonomy. However, many studies have sought to exclude complex local-world patterns of taxi demand from the historical taxi demand threshold, effectively ignoring the underlying effects of regional activity and effectively mobilizing long-term cycles. In the article by Cao et al. [8], they note two significant observations; one is that the pattern of taxi demand varies significantly between different active areas, and the other is the demand for taxis following dynamic daily and weekly patterns. To address these two issues, they proposed a new bidirectional encoder representation-based deep spatial-temporal network that captures locations of interest that define regional functions and include multiple local and indicates complex local-temporal relationships with global features. Bidirectional encoder representations-based deep spatial-temporal network' points of interest have introduced a time-space adaptive module to capture taxi demand's complex time-space pattern and dynamic time phases. Points of interest have implemented a functional agreement implantation module in all regions. To their knowledge, this is the first time the proposed architecture has been used to determine the types of taxi claims, and this is the first time they have considered the practical similarities. Their research results with the New York City Real Traffic Database show that the proposed method implements more complex methods and their suggested model is much better than other methods.

In the article by Xu et al. [28], they suggest a sequential learning model that can estimate the demand for taxis in different areas of the city depending upon current demand and other relevant information. It is essential to consider the advanced information here as future taxi applications will be linked to past activities. For example, anyone requesting a taxi at the mall can request a taxi to reach home within

hours. They use LSTM, an excellent sequential learning system, to store relevant information for future use. They evaluate their perspective by dividing the city into smaller areas and estimating the needs of each area within the New York City Data Application Database. Furthermore, they have shown that this system surpasses other predictive methods, such as future neural networks. In addition, they show how additional relevant information such as time, duration, and reduction affect the results. In the article by Yu et al. [29], a framework is proposed to suggest the needs of taxi passengers. They consider temporal, spatial, and external dependencies. The proposed deep learning framework integrates an updated density-based spatial clustering algorithm with noise and conditional generative adversarial network models. More specifically, the updated model is applied to the road network to create multiple subnetworks that consider the local relevance of taxi pick-up events. Comparative results show that the proposed model outperforms all other methods. It is recommended that more data be added to test future research models and that more information be added to improve prediction performance.

Taxi origin and destination flow forecasts for any city play an important role in passenger travel needs and taxi management and scheduling. However, complex local dependence and temporal mobility make this problem difficult. In the article by Duan et al. [30], a predictive model of a hybrid deep neural network based on convoluted LSTM was proposed. The underlying relationship between travel time and origin and destination flow was investigated to improve the prediction accuracy and integrated into the prediction model as input. Actual taxi data was used to fully validate the experiment's proposed model and forecasting system. Taxi demand forecasting is essential for making decisions on online taxi application platforms. The article by Zhang et al. [31] designed and explored how mutual variations can be used to improve grouping and area forecasting. First, a taxi zone grouping algorithm was developed based on the theory of grouping in pairs, which considers the relationships between the different taxi zones. Then, group-level and global prediction modules were developed to achieve internal and intercluster features, respectively. Finally, a multilevel recurrent neural network model was proposed to combine the two modules.

**2.4. Electric Vehicle Charging Station Adoption.** The construction of a charging station in the traffic network is a significant step toward the development of intelligent vehicle systems in urban areas [32–34]. Due to their high energy efficiency and low emissions, electric vehicles have become an attractive means of transportation to develop clean mobility systems. Infrastructure-based automation needs recharging facilities to meet the growing demand. The planning process is, of course, responsible for their impact on the power grid [35]. Chen et al. [36] first developed a new user balance model that described the distribution of road network balances along the charging lane. The battery re-charge plan specified which charging lane to use, how long to charge, and which electric vehicle to drive. They developed

charging lane applications under established user balance conditions as a math program with more complementary boundaries. The network balance and the design model were solved by efficient solution algorithms and illustrated with numerical examples. The Flow Refueling Location Model (FRLM) proposed by Kuby and Lim [37] is a flow interaction model that identifies locations on a network to maximize essential destination flow. Due to the limited driving distance of the vehicle, the network does not form a set of vertical dominance seats. The work by Miralinaghi et al. [38] provides a two-tiered mathematical model for understanding the decision-making process of transport companies and passengers. Under the structure, EV charge is a robust theoretical basis for network design. The design problem is solved using a functional set algorithm. The study results could serve as a guide for metropolitan transport companies to build capacity for specific locations and EVs, thus reducing long-term emissions. In another article by Miralinaghi et al. [39], they considered the problem of identifying charging stations in the transport network through mathematical programming. The proposed model applies to various alternative fuels and is particularly suitable for hydrogen fuels. They applied two well-known solution algorithms, branch-and-bound algorithm and Lagrangian relaxation algorithm, to solve the problem.

### 3. Methodology

We obtained the data from an electric kickboard provider, and then we used the  $k$ -means for clustering and bagging regressor for the final demand prediction.  $k$ -means clustering is a type of unsupervised learning [40].  $k$ -means is used with 4 clusters. Bagging is a method of taking multiple samples and aggregating the results based on them, and each sample independently predicts the result [41]. The bagging regressor is used with three different models. Extreme Gradient Boosting (XGBoost), Extra Trees, and Random Forest are used for the base layer, and the Extra Trees model is used for the metalayer.

Extreme Gradient Boosting is a Gradient Boosting technique that reflects the weight of the result from the first sample to the following sample as opposed to bagging, where each sample independently predicts the result [42]. It continues to learn the weights of the results from the previous sample to affect the following sample as well. XGBoost has a faster learning rate and better model performance than other models based on Gradient Boosting. Gradient Boosting concentrates only on the training data results, and overfitting easily occurs. XGBoost can prevent overfitting by adjusting the hyperparameter values provided by the programmers by setting the desired learning method.

Random Forest is an algorithm that makes decisions through multiple decision trees, and it was created on the assumption that numerous ordinary algorithms solve problems better than one smart algorithm [43]. As a learning method, the final prediction value is determined by collecting the results determined by several trees. Random Forest is a representative bagging method and represents a voting method. Voting is a method of finally predicting the

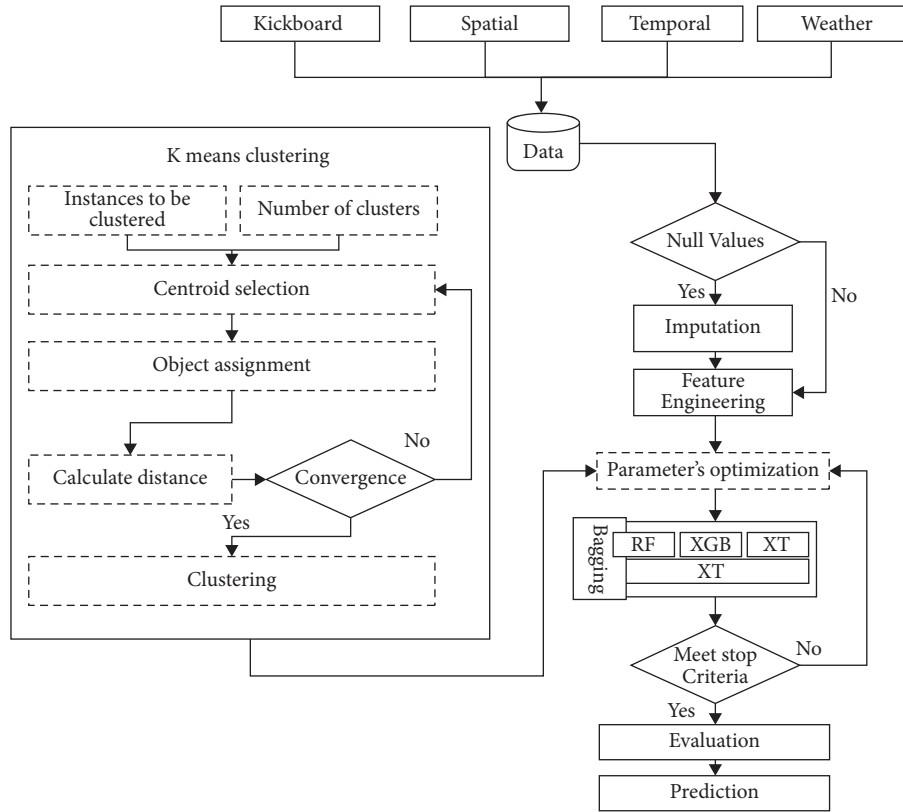


FIGURE 3: Flowchart of the proposed methodology.

highest value using the results of several samples [18]. Instead of using a single decision tree, learning is performed as many as the number of decision trees the programmer sets, and the results are collected. The most mentioned result is used as the final prediction value.

Extra Trees regressor learns in a similar way to Random Forest. However, Random Forest uses all feature values to produce results, whereas Extra Trees selects a method in which multiple decision trees randomly select features to produce optimal results [44]. Therefore, the learning speed of Extra Trees randomly selecting some features is faster than Random Forest using all feature values. Hence, we have used it as a metalearner.

Figure 3 shows the flow diagram of the proposed methodology. It starts with data aggregation. We combined the data from different sources such as kickboard, spatial, temporal, and weather data. Kickboard-related information consists of rent date, rent number, and sector information. Spatial information consisted of the  $x$  position and  $y$  position of the kickboard. Temporal information consisted of rent day, year, and month whereas weather data consisted of temperature, rain, humidity, and insolation. We checked whether there was any null value; if not, we performed feature engineering. If there were some null values, they were imputed using the mean. Feature engineering consisted of creating new features from existing parameters, such as extracting day, month, and week information from timestamps. The selected parameters were passed to the  $k$ -means clustering module where the instance was clustered, and the

number of clusters was passed for centroid selection. We chose four clusters. The next step was to assign objects to the nearest distance and then calculate the distance. If convergence was achieved, then the final clustering step would be finished; otherwise, it would start again from centroid selection. The selected clusters were transferred for the optimal parameter selection. These optimal parameters were forwarded to the bagging regressor model, consisting of three base models and one metamodel. Random Forest, XGBoost, and Extra Trees regressor were used as base models, whereas Extra Trees was used as a metamodel to obtain the final prediction. The model was evaluated using different evaluation metrics such as  $R$ -squared, root mean square error, and Kolmogorov–Smirnov test, and then a prediction was made.

Figure 4 shows the structure of the proposed methodology. We have collected the electric kickboard data from the local kickboard company of Jeju Island, South Korea. Jeju Island is a famous island for tourism in South Korea. Tourists use electric kickboards to move from one place to another. We got the vehicle, spatial, and temporal information from them. Then, we got the weather information from Korea Metrological Department. We performed exploratory data analysis to understand the nature of the data. This data was preprocessed using different techniques, such as removing outliers and feature extraction. This preprocessed data was forwarded to the  $k$ -means clustering algorithm, where we found different clusters in Jeju Island. These clusters were passed to the bagging ensemble model



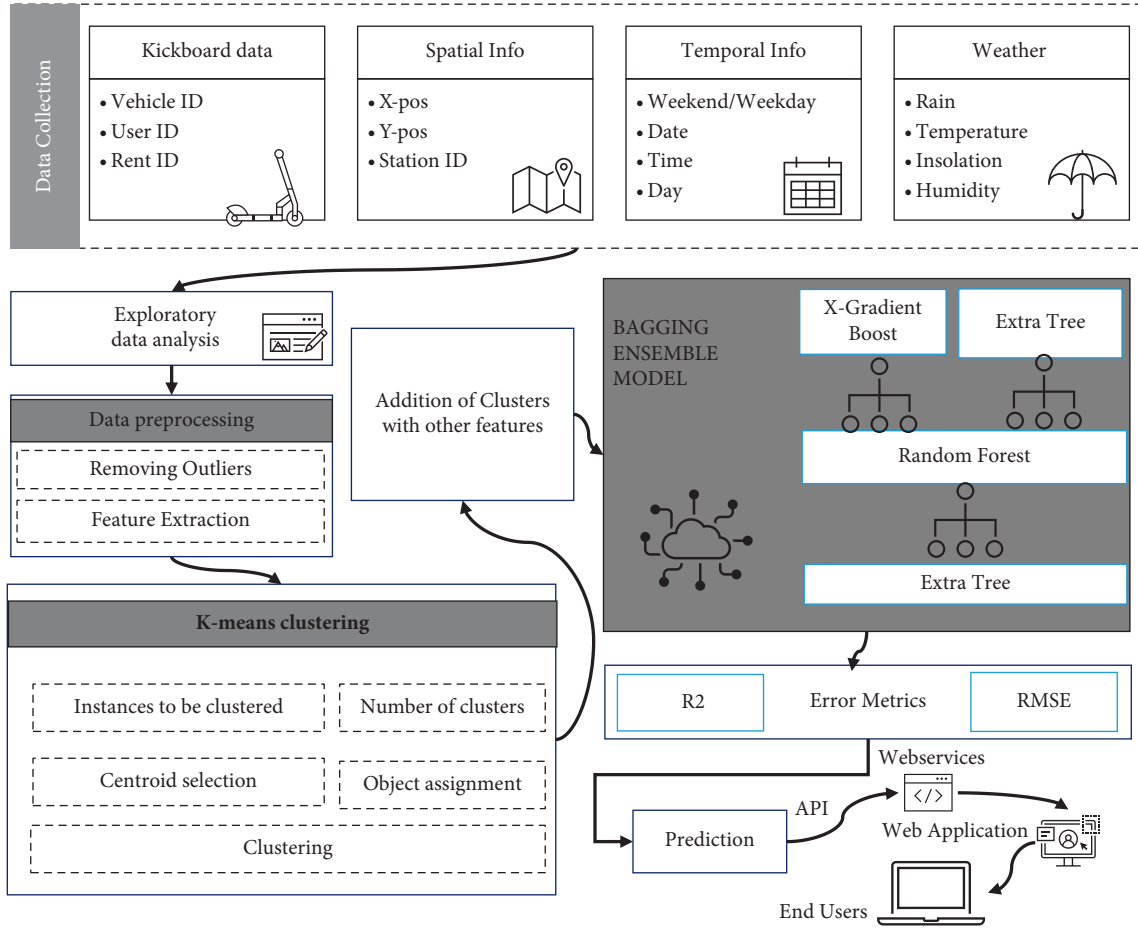


FIGURE 4: Structure of the proposed methodology.

and other spatiotemporal and vehicle features. This ensemble model consists of two layers [45]. Layer 0, or the base layer, consists of an Extreme Gradient Boosting, Random Forest, and Extra Trees model. Layer 1 or metalayer consists of an Extra Trees model. We used different evaluation metrics such as *R*-squared score and root means square error (RMSE) to evaluate our proposed approach. The final prediction was forwarded to the web application using web services. End users could see the future prediction about electric kickboard demand for specific locations.

#### 4. Data Analysis

In recent years, the electric scooter, kickboards, and bikes in many cities across the world have provided an excellent opportunity to reduce short-distance driving [46]. The data used for the study was the demand data of electric mobility (electric kickboard) service company that started services on Jeju Island in April 2019. Data from EV Pass company, a company that provides electric mobility services on Jeju Island, was used. The total number of data instances is service users' number of use cases during the data collection period. The collection period is 717 days, from April 16, 2019, to June 11, 2021. This study predicts the daily demand for electric mobility by grouping the demand for electric mobility by day. In addition, we imported weather data from

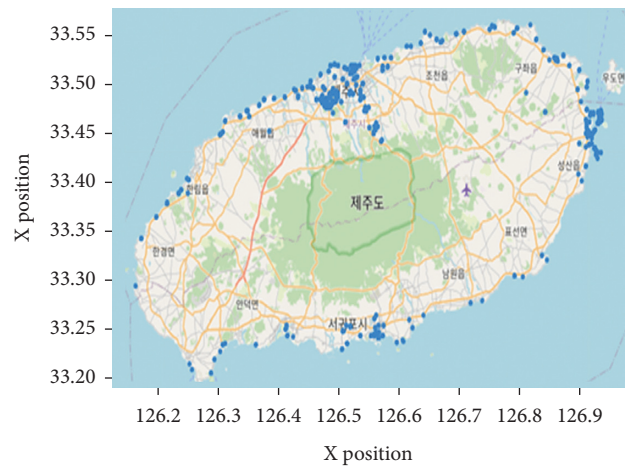


FIGURE 5: Rental location according to latitude and longitude.

the Korea Meteorological Administration [47] and added external factors that affect the use of electric mobility, such as daily average temperature and precipitation, and dividing weekends and weekdays.

Figure 5 shows rental locations according to latitude and longitude. The *x*-axis represents the *x* position or latitude, and the *y*-axis shows the *y* position or longitude.

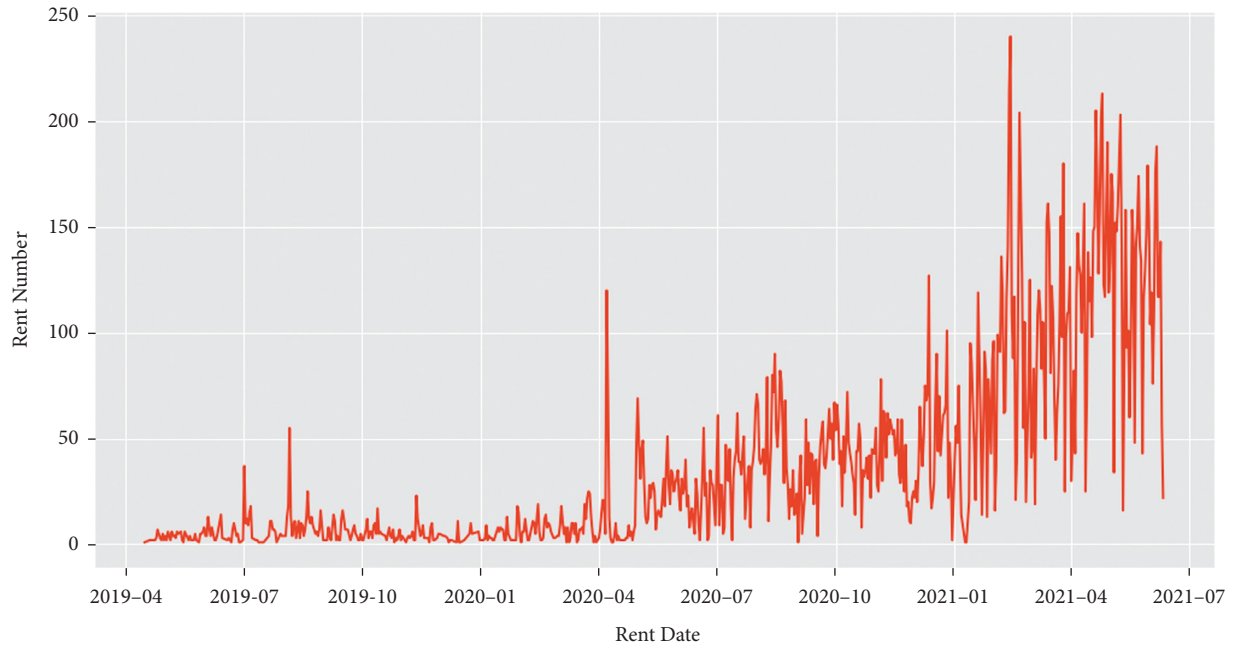


FIGURE 6: Daily demand for electric kickboards.

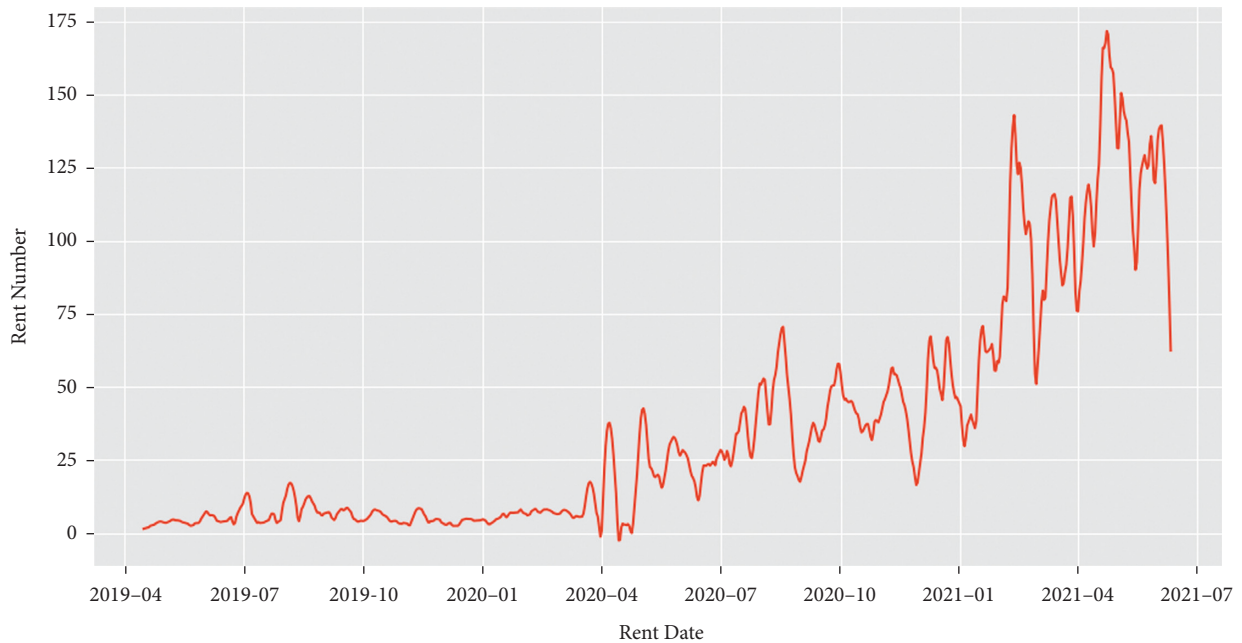


FIGURE 7: Daily demand for electric kickboards after smoothing the data.

Figure 6 shows daily demand for electric kickboards. The  $x$ -axis represents the date, and the  $y$ -axis represents the total daily demand for that day. Demand, which was low at the beginning of the service, increased after a certain period of time. This can be thought of as a demand phenomenon for startups in general. In addition, demand was not constant and showed large differences from day to day. This study was conducted by stabilizing these unstable data through

smoothing. The window size for data smoothing was set to 11, and the window size standardizes the data by grouping the day and the days before and after it. The window size of 11 means the midday and five days before and after, and Figure 7 shows the data smoothing result.

Figure 8 explains the effect of the holiday, weekend, and weekday on rental kickboard demand. The  $x$ -axis represents the feature name, and the  $y$ -axis shows the average rent



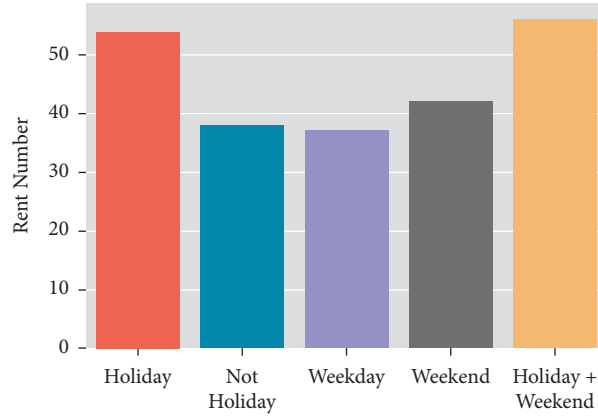


FIGURE 8: Effect of the holiday, weekend, and weekday on rental kickboard demand.

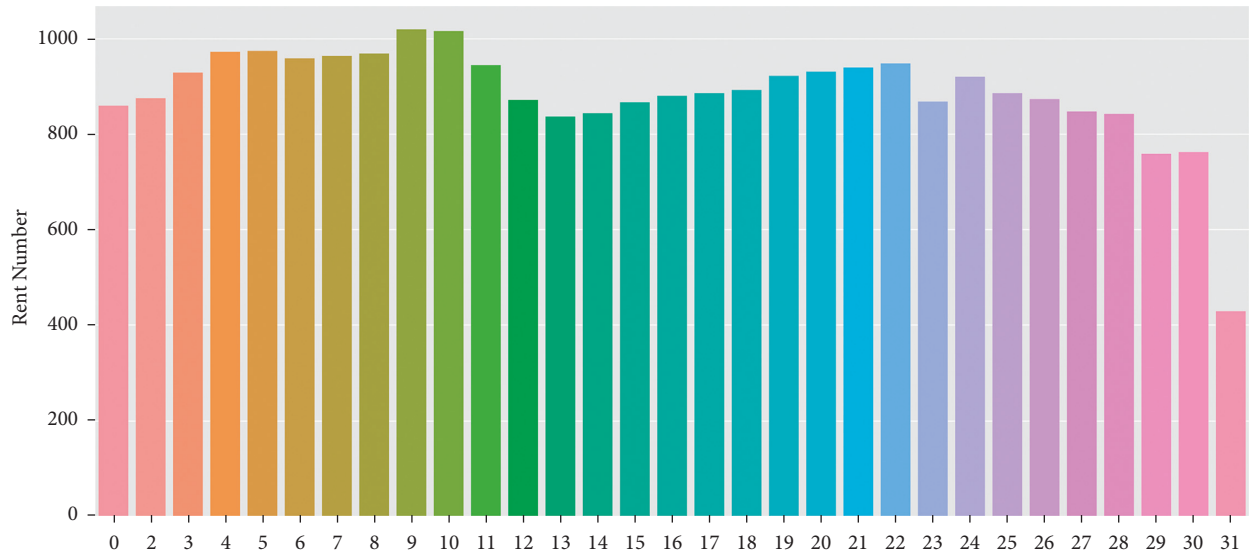


FIGURE 9: Total Demand on each day of a month.

TABLE 2: Simulation environment.

Sr#	Property	Specification
1	Programming language	Python 3.7.6
2	Operating system	Ubuntu 18.04
3	Browser	Google Chrome
4	Framework	Jupyter Notebook
5	CPU	Intel Core i7 CPU @ 1.80 GHZ
6	Memory	16 GB
7	Graphic card	Radeon 540

number for that day. The holiday accounts for the most rental kickboards, whereas weekdays account for the least numbers if there is a holiday and weekend with the maximum number of rental kickboards.

Figure 9 explains demand on each day of a month. The  $x$ -axis represents the day number of the month, and the  $y$ -axis shows the total rent number on that specific day.

## 5. Results

This section covers the experimental results achieved using our presented approach. We have also compared our presented model with state-of-the-art algorithms. We have used the Jupyter notebook on Ubuntu 18.04 for coding in Python 3.7.6. Table 2 summarize the simulation environment used

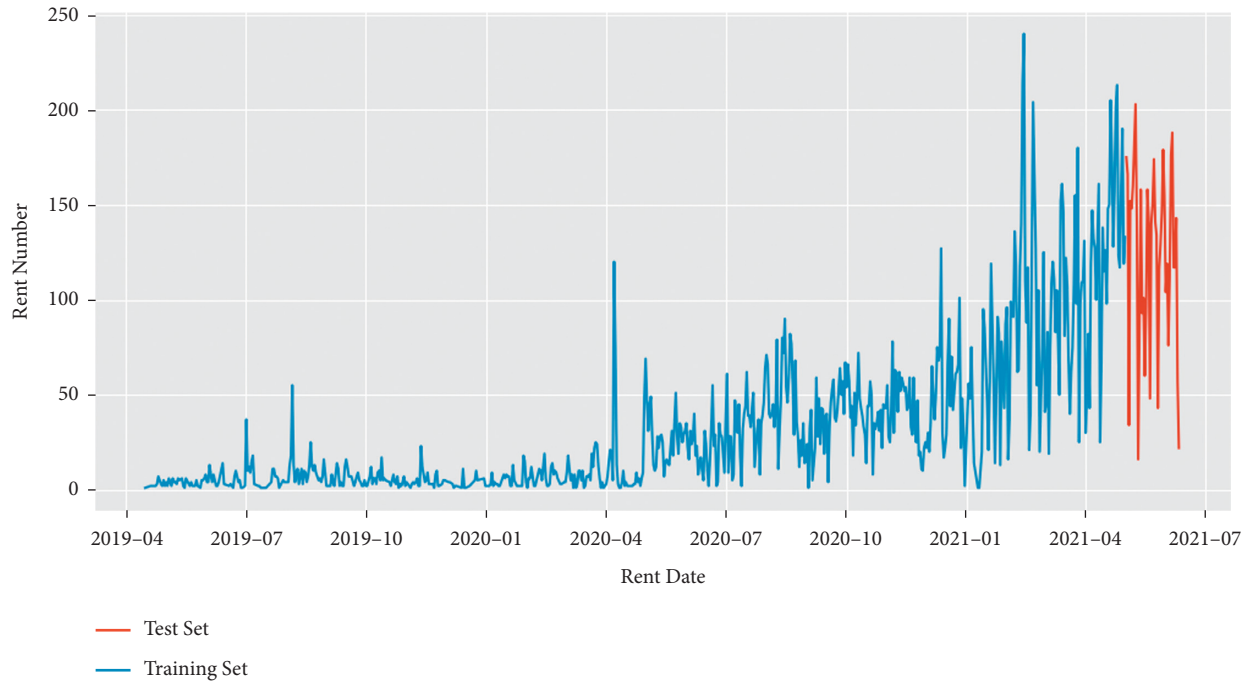


FIGURE 10: Training and testing data distribution.

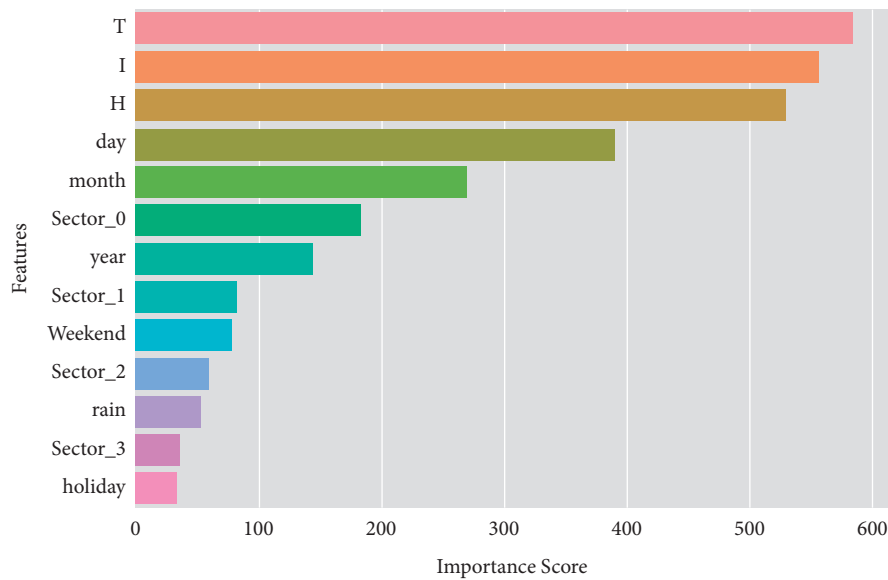


FIGURE 11: Feature importance graph.

for this research. We have used 80 percent of data for model training and 20 percent for result validation. Figure 10 reflects the training and test data distribution. The  $x$ -axis represents the date, and the  $y$ -axis shows the rent count. The blue line shows the training set, and the black line shows the testing dataset.

Feature importance is also commonly referred to as the feature selection method. When supervised learning makes predictions through learned data, it is a numerical expression of the effect of each feature on the result. In the time

series data sorted by day, the feature importance was higher as the feature had a clear pattern and subdivided numerical values among the information representing each day. Figure 11 depicts feature importance graph.  $T$  represents temperature,  $I$  represents insulation, and  $H$  represents humidity. Sector 0 is the first sector, and sector 3 is the fourth sector. Other features are day, month, year, weekend, rain, and holiday. It is observed that temperature, insulation, and humidity impact the final prediction. The  $x$ -axis represents the feature importance score, and the  $y$ -axis shows the name

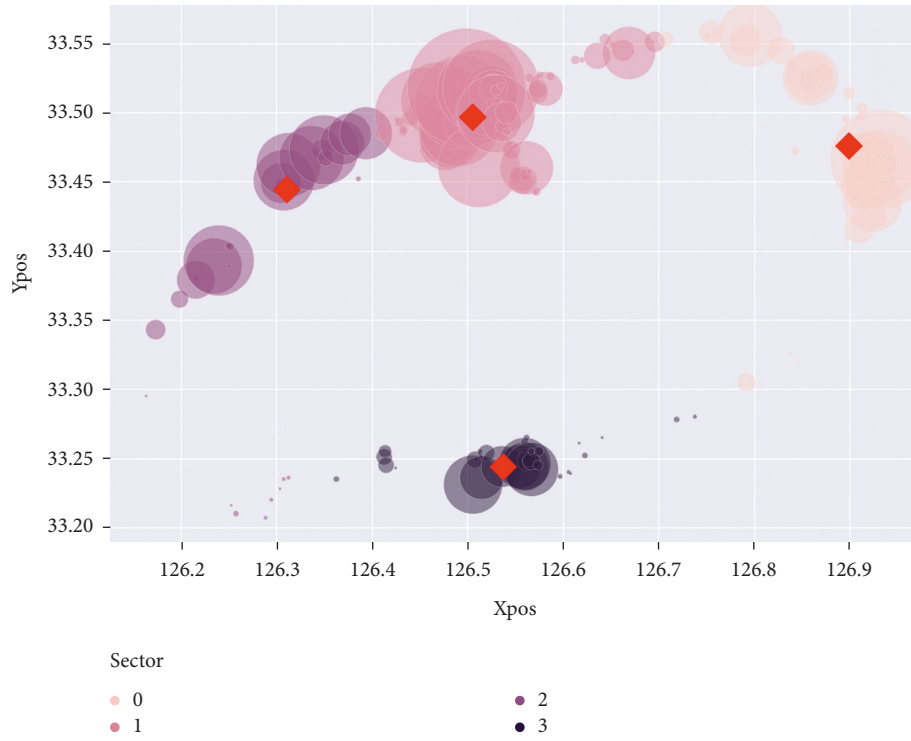


FIGURE 12: Clustering with four sectors.

of the features. The feature importance is calculated using the importance score where temperature, insulation, and humidity have more importance than rain in the final prediction.

In segmenting regions, the division results based on latitude and longitude are also compared with the results of clustering division using the  $k$ -means algorithm. Figure 12 displays clustering with 4 clusters. Apply  $k$ -means clustering with 4 clusters. Red spots represent the centroid of the cluster. Sectors are numbered 0, 1, 2, and 3. The  $x$ -axis represents the  $x$  position or latitude, and the  $y$ -axis shows the  $y$  position or longitude.

Figure 13 explains sector-wise demand, where sector 1 is the hotspot for rental kickboards and sector 3 has minimum kickboard demand. Sector 1 represents the Jeju-si district, and sector 3 represents the Seogwipo-si district. The population of Jeju-si is around 492 thousand, while that of Seogwipo-si is 179 thousand. Moreover, the total population of Jeju-si stands at around 671 thousand [48]. The difference in population is also a reason for more demand in sector 1. The  $x$ -axis represents the sector number, and the  $y$ -axis shows the total rent number in that specific sector. The figure shows the difference in the size of the dots according to the total number of rentals at each rental office on Jeju Island during the data collection period. It can be seen that the number of electric mobility rentals near Jeju Airport and famous tourist destinations such as Aewol and Seongsan is generally large. The demand for electric scooters is usually high at tourist attractions. Jeju Island consists of many tourist attractions such as Hallasan Mountain Natural

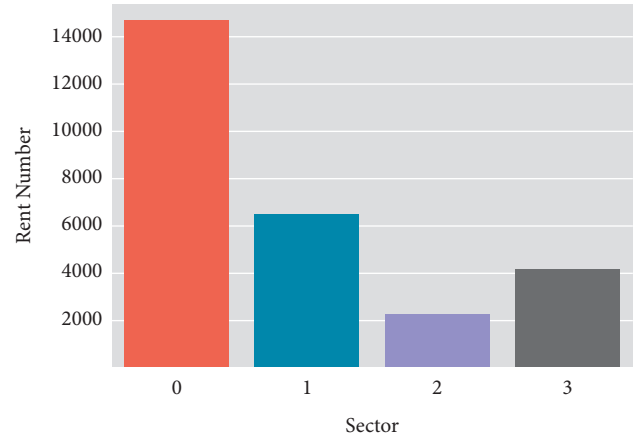


FIGURE 13: Sector-wise demand.

Reserve, Geomunoreum Lava Tube System, and Seongsan Ilchulbong Tuff Cone.

Figure 14 shows prediction results. The light blue line shows the actual value, and the black line shows the predicted value. This graph shows the result of the proposed approach for the test data. We have used test data from May 1, 2021, to June 11, 2021. The  $x$ -axis represents the date, and the  $y$ -axis shows the rent count.

Root mean square error (RMSE) is defined as the square root of mean square error (MSE) [49]. It is used to measure the difference from the actual values of the predicted values. The formula for calculating RMSE is given in (1), where  $n$  is the total number of observations,  $y_{ob}$  represents the

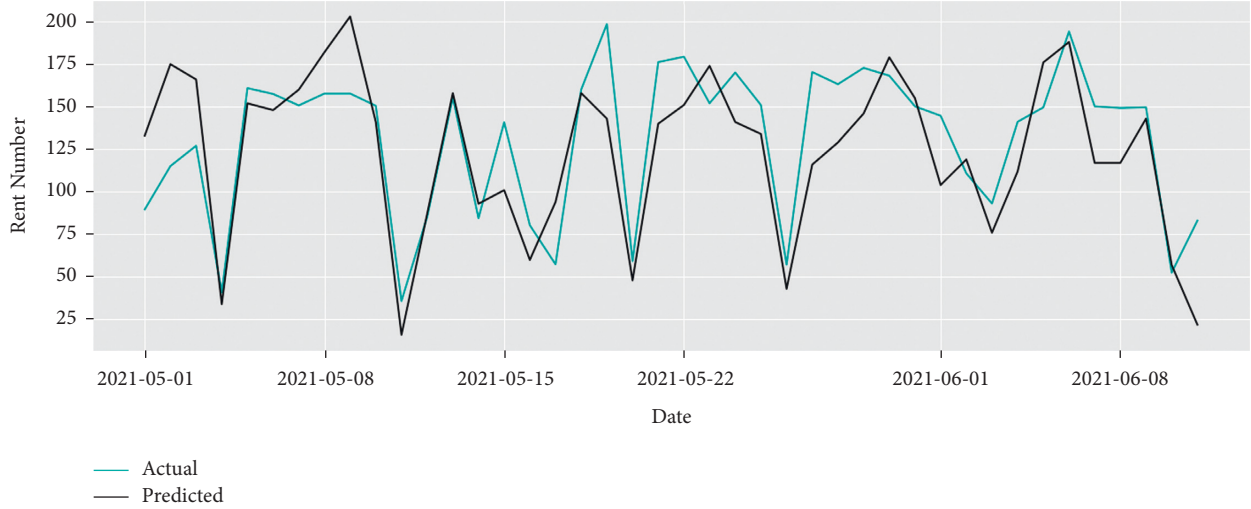


FIGURE 14: Prediction result of the proposed approach.

TABLE 3: Comparison of the proposed approach with different individual models.

Sr#	Model	$R^2$ score	RMSE
1	CatBoost	78.21	46.78
2	Extra Trees	91.67	21.23
3	XGBoost	89.64	29.91
4	LGBM	84.48	36.68
5	Random Forest	87.12	27.61
6	Gradient Boosting	68.17	39.31
7	Proposed	93.42	24.6

TABLE 4: Comparison of the proposed approach with different combinations.

Hybrid model combinations		$R^2$ score
Base layer	Metalayer	
CatBoost + Extra Trees + LGBM	CatBoost	67.67
CatBoost + Extra Trees + Random Forest	Extra Trees	72.57
CatBoost + Extra Trees + XGBoost	XGBoost	70.47
CatBoost + LGBM + Random Forest	Random Forest	79.24
Extra Trees + XGBoost + Random Forest	Random Forest	87.32
Extra Trees + XGBoost + Random Forest	XGBoost	89.05
Extra Trees + XGBoost + Random Forest	Extra Trees	93.42

observed actual value, and  $y_{es}$  represents the estimated (predicted) value. We have achieved an RMSE of 24.67 using our proposed approach. Table 3 shows a comparison of the proposed approach with different individual models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{ob} - y_{es})^2}{n}}. \quad (1)$$

The  $R^2$  score or regression score is a statistical measure that is defined as a set coefficients that involves observed and predicted values [50]. A regression score is used to estimate how well the reaction model works.  $R^2$  is an indication of the good performance achieved by the reaction model if it is near score 1 and bad performance if the value is near to zero.

The  $R^2$  score is calculated based on (2), where  $n$  is the total number of observations,  $y_{ob}$  represents the observed actual value, and  $y_{es}$  represents the estimated (predicted) value.

$$R^2 \text{ score} = 1 - \frac{\sum (y_{ob} - y_{es})^2}{\sum (y_{ob} - \bar{y}_{es})^2}. \quad (2)$$

We have achieved  $R^2$  of 93.42 using our proposed approach. We have compared our model with different other combinations of bagging regressors, as shown in Table 4 which displays the models used in the base layer and the metalayer.

The Kolmogorov–Smirnov goodness of fit test (K–S test) compares the data with a known distribution and shows if

TABLE 5: Kolmogorov–Smirnov goodness of fit test.

Feature	Statistics	<i>P</i> value
Sector 1	0.052	1.00
Sector 2	0.091	0.68
Sector 3	0.119	0.44
Sector 4	0.013	1.00

they have the same distribution. We performed the K–S test on our four sectors. The results of the K–S test are represented in Table 5.

The key element in the proposed model is the bagging ensemble method. This method allows one to take advantage of different models and combine them into one model. Another critical step proposed in this article is to cluster the demand according to regions—this helps improve the accuracy. The major drawback of other forecasting models that result in less accurate forecasting is the lack of ensemble technique.

## 6. Conclusions

The demand for electric mobility is increasing, especially in tourist attractions. Machine learning can help accurately predict the electric mobility demand in areas where companies struggle to meet the demand at the proper location. This article performs classification-aided bagging regressor prediction of supervised learning using a small amount of data and unstable data with a significant difference in demand. We have used the data from a local electric kickboard provider on Jeju Island, South Korea. The company provides electric kickboards for rent to tourists and residents. Data smoothing stabilized the irregular pattern between data while maintaining the overall demand pattern. Data with similar characteristics were grouped using clustering, and data with different characteristics were separated to predict demand. We have utilized the *k*-means algorithm for clustering and bagging ensemble model for final prediction. The bagging ensemble model consists of XGBoost, Extra Trees, and Random Forest algorithms. We have used an Extra Trees model as a metalearner for the proposed model. Through this, the forecasting results significantly increased, and the demand forecasting accuracy results after the regional division of Jeju Island were also enhanced. We have achieved  $R^2$  of 93.42 using our proposed approach. The results of this study can be helpful for the electric mobility providers who want to predict the demand at a specific city location accurately. In the future, genetic algorithm can be used for feature optimization and hyperparameter optimization.

## Abbreviations

ML:	Machine learning
LSTM:	Long short-term memory
GARCH:	Generalized autoregressive conditional heteroskedasticity
SARIMA:	Seasonal autoregressive moving average

BDSTN:	Bidirectional deep spatial-temporal network
HDBSCAN:	Hierarchical density-based spatial clustering of applications with noise
H-ConvLSTM:	Hexagonal convolution operation LSTM
DSTGNN:	Dynamic spatiotemporal graph neural networks
GHG:	Greenhouse gas emissions
EV:	Electric vehicle
XGboost:	Extreme gradient boost
RMSE:	Root mean square error
K-S Test:	Kolmogorov-smirnov test

## Data Availability

The data used to support this study are available from the corresponding authors upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the 2021 scientific promotion program funded by Jeju National University.

## References

- [1] E. Howe and B. Bock, “Global scootersharing market report 2018,” *InnoZ-Innovation Centre for Mobility and Societal Change*, (InnoZ) GmbH, Berlin, Germany, 2018.
- [2] EVPASS, “Rental Products and Fees,” 2022, <https://www.evpass.co.kr/web/pricetable.php?product=>.
- [3] S. Kim and K. Kim, “A study on the startup growth stage in Korea,” *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, vol. 15, no. 2, pp. 127–135, 2020.
- [4] G. Zarazua de Rubens, L. Noel, J. Kester, and B. K. Sovacool, “The market case for electric mobility: investigating electric vehicle business models for mass adoption,” *Energy*, vol. 194, Article ID 116841, 2020.
- [5] J. Degele, A. Gorr, K. Haas, D. Kormann, S. Krauss, and P. Lipinski, “Identifying E-scooter sharing customer segments using clustering,” in *Proceedings of the 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1–8, IEEE, Stuttgart, Germany, June 2018.
- [6] Y. Lai, K. Zhang, J. Lin, F. Yang, and Y. Fan, “Taxi demand prediction with LSTM-based combination model,” in *Proceedings of the 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pp. 944–950, IEEE, Xiamen, China, December 2019.
- [7] N. Saum, S. Sugiura, and M. Piantanakulchai, “Short-term demand and volatility prediction of shared micro-mobility: a case study of e-scooter in thammasat university,” in *Proceedings of the 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS)*, pp. 27–32, IEEE, Delft, Netherlands, November 2020.
- [8] D. Cao, K. Zeng, J. Wang et al., “BERT-based deep spatial-temporal network for taxi demand prediction,” *IEEE*

- Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9442–9454, 2022.
- [9] M. Heumann, T. Kraschewski, T. Brauner, L. Tilch, and M. H. Breitner, “A spatiotemporal study and location-specific trip pattern categorization of shared E-scooter usage,” *Sustainability*, vol. 13, no. 22, Article ID 12527, 2021.
  - [10] S. Kim, S. Choo, G. Lee, and S. Kim, “Predicting demand for shared E-scooter using community structure and deep learning method,” *Sustainability*, vol. 14, no. 5, p. 2564, 2022.
  - [11] F. Zwick and K. W. Axhausen, “Ride-pooling demand prediction: a spatiotemporal assessment in Germany,” *Journal of Transport Geography*, vol. 100, Article ID 103307, 2022.
  - [12] K. Liu, Z. Chen, T. Yamamoto, and L. Tuo, “Exploring the Impact of Spatiotemporal Granularity on the Demand Prediction of Dynamic Ride-Hailing,” 2022, <https://arxiv.org/abs/2203.10301>.
  - [13] F. Huang, P. Yi, J. Wang, M. Li, J. Peng, and X. Xiong, “A Dynamical Spatial-Temporal Graph Neural Network for Traffic Demand Prediction,” *Information Sciences*, vol. 594, 2022.
  - [14] Y. Kazancoglu, M. Ozbiltekin-Pala, and Y. D. Ozkan-Ozen, “Prediction and evaluation of greenhouse gas emissions for sustainable road transport within Europe,” *Sustainable Cities and Society*, vol. 70, Article ID 102924, 2021.
  - [15] G. Xu, Y. Lv, H. Sun, J. Wu, and Z. Yang, “Mobility and evaluation of intercity freight CO<sub>2</sub> emissions in an urban agglomeration,” *Transportation Research Part D: Transport and Environment*, vol. 91, Article ID 102674, 2021.
  - [16] P. W. Khan and Y. C. Byun, “Smart contract centric inference engine for intelligent electric vehicle transportation system,” *Sensors*, vol. 20, no. 15, p. 4252, 2020.
  - [17] Z. Yi, X. C. Liu, R. Wei, X. Chen, and J. Dai, “Electric vehicle charging demand forecasting using deep learning model,” *Journal of Intelligent Transportation Systems*, vol. 2021, Article ID 1966627, 1–14 pages, 2021.
  - [18] N. Zuniga-Garcia, N. R. Juri, K. A. Perrine, and R. B. Machemehl, “E-Scooters in Urban Infrastructure: Understanding Sidewalk, Bike Lane, and Roadway Usage from Trajectory Data,” *Case Studies on Transport Policy*, vol. 9, no. 3, 2021.
  - [19] N. Davies, L. Blazejewski, and G. Sherriff, “The rise of micromobilities at tourism destinations,” *Journal of Tourism Futures*, vol. 6, no. 3, pp. 209–212, 2020.
  - [20] S. Bai and J. Jiao, “Dockless E-scooter usage patterns and urban built Environments: a comparison study of Austin, TX, and Minneapolis, MN,” *Travel behaviour and society*, vol. 20, pp. 264–272, 2020.
  - [21] Y. Feng, D. Zhong, P. Sun et al., “Micromobility in smart cities: a closer look at shared dockless E-scooters via big social data,” in *Proceedings of the ICC 2021-IEEE International Conference on Communications*, pp. 1–6, IEEE, Montreal, QC, Canada, June 2021.
  - [22] M. Kolaković-Bojović and A. Paraušić, “Electric scooters-urban security challenge or moral panic issue,” *Temenos*, vol. 43, no. 4, p. 1045, 2020.
  - [23] Q. Ma, S. Li, H. Zhang, Y. Yuan, and L. Yang, “Robust optimal predictive control for real-time bus regulation strategy with passenger demand uncertainties in urban rapid transit,” *Transportation Research Part C: Emerging Technologies*, vol. 127, Article ID 103086, 2021.
  - [24] A. Achar, D. Bharathi, B. A. Kumar, and L. Vanajakshi, “Bus arrival time prediction: a spatial Kalman filter approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1298–1307, 2020.
  - [25] F. Tianheng, Y. Lin, G. Qing, H. Yanqing, Y. Ting, and Y. Bin, “A supervisory control strategy for plug-in hybrid electric vehicles based on energy demand prediction and route preview,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 1691–1700, 2015.
  - [26] L. Liu, Z. Qiu, G. Li, Q. Wang, W. Ouyang, and L. Lin, “Contextualized spatial-temporal network for taxi origin-destination demand prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3875–3887, 2019.
  - [27] C. Zhang, F. Zhu, X. Wang, L. Sun, H. Tang, and Y. Lv, “Taxi Demand Prediction Using Parallel Multi-Task Learning Model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, 2020.
  - [28] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, “Real-time prediction of taxi demand using recurrent neural networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2018.
  - [29] H. Yu, X. Chen, Z. Li et al., “Taxi-based mobility demand formulation and prediction using conditional generative adversarial network-driven learning approaches,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3888–3899, 2019.
  - [30] Z. Duan, K. Zhang, Z. Chen et al., “Prediction of city-scale dynamic taxi origin-destination flows using a hybrid deep neural network combined with travel time,” *IEEE Access*, vol. 7, pp. 127816–127832, 2019.
  - [31] C. Zhang, F. Zhu, Y. Lv, P. Ye, and F. Y. Wang, “MLRNN: Taxi Demand Prediction Based on Multi-Level Deep Learning and Regional Heterogeneity Analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, 2021.
  - [32] M. Miralinaghi, Y. Lou, B. B. Keskin, A. Zarrinmehr, and R. Shabanpour, “Refueling station location problem with traffic deviation considering route choice and demand uncertainty,” *International Journal of Hydrogen Energy*, vol. 42, no. 5, pp. 3335–3351, 2017.
  - [33] P. W. Khan and Y. C. Byun, “Blockchain-based peer-to-peer energy trading and charging payment system for electric vehicles,” *Sustainability*, vol. 13, no. 14, p. 7962, 2021.
  - [34] P. W. Khan and Y. C. Byun, “Secure transactions management using blockchain as a service software for the internet of things,” in *Software Engineering in IoT, Big Data, Cloud and Mobile Computing*, pp. 117–128, Springer, 2021.
  - [35] B. Azin, X. T. Yang, N. Marković, and M. Liu, “Infrastructure enabled and electrified automation: charging facility planning for cleaner smart mobility,” *Transportation Research Part D: Transport and Environment*, vol. 101, Article ID 103079, 2021.
  - [36] Z. Chen, F. He, and Y. Yin, “Optimal deployment of charging lanes for electric vehicles in transportation networks,” *Transportation Research Part B: Methodological*, vol. 91, pp. 344–365, 2016.
  - [37] M. Kuby and S. Lim, “Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs,” *Networks and Spatial Economics*, vol. 7, no. 2, pp. 129–152, 2007.
  - [38] M. Miralinaghi, G. H. de Almeida Correia, S. E. Seilabi, and S. Labi, “Designing a network of electric charging stations to mitigate vehicle emissions,” in *Proceedings of the 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS)*, pp. 95–100, IEEE, Delft, Netherlands, November 2020.
  - [39] M. Miralinaghi, B. B. Keskin, Y. Lou, and A. M. Roshandeh, “Capacitated refueling station location problem with traffic deviations over multiple time periods,” *Networks and Spatial Economics*, vol. 17, no. 1, pp. 129–151, 2017.



- [40] P. Tavallali, P. Tavallali, and M. Singhal, "K-means tree: an optimal clustering tree for unsupervised learning," *The Journal of Supercomputing*, vol. 77, no. 5, pp. 5239–5266, 2021.
- [41] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, "Ensemble boosting and bagging based machine learning models for groundwater potential prediction," *Water Resources Management*, vol. 35, no. 1, pp. 23–37, 2021.
- [42] K. Jung, J. Park, S. Son, and S. Ahn, "Position prediction of wireless charging electric vehicle for auto parking using extreme gradient boost algorithm," in *Proceedings of the 2020 IEEE Wireless Power Transfer Conference (WPTC)*, pp. 439–442, IEEE, Seoul, Korea (South), November 2020.
- [43] Y. Zhan, Y. Luo, X. Deng, M. L. Grieneisen, M. Zhang, and B. Di, "Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment," *Environmental Pollution*, vol. 233, pp. 464–473, 2018.
- [44] D. Meddage, I. U. Ekanayake, A. Weerasuriya, and C. Lewangamage, "Tree-based regression models for predicting external wind pressure of a building with an unconventional configuration," in *Proceedings of the 2021 Moratuwa Engineering Research Conference (MERCon)*, pp. 257–262, IEEE, Moratuwa, Sri Lanka, July 2021.
- [45] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," *Energy*, vol. 214, Article ID 118874, 2021.
- [46] J. Oh and J. Kim, "Where to ride? An explorative study to investigate potential risk factors of personal mobility accidents," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, p. 965, 2021.
- [47] K. M. Administration, 2022, <https://www.kma.go.kr/eng/index.jsp>.
- [48] P. Statista, "distribution of Jeju," 2021, <https://www.statista.com/statistics/756061/south-korea-population-distribution-of-jeju-by-district/>.
- [49] S. Sakr, R. Elshaw, A. M. Ahmed et al., "Comparison of machine learning techniques to predict all-cause mortality using fitness data: the Henry ford exercise testing (FIT) project," *BMC Medical Informatics and Decision Making*, vol. 17, no. 1, p. 174, 2017.
- [50] P. W. Khan, Y. C. Byun, and S. J. Lee, "Optimal photovoltaic panel direction and tilt angle prediction using stacking ensemble learning," *Frontiers in Energy Research*, p. 10, 2022.