proceedings

MDPI

Proceeding Paper

Enhancing Youbike Redistribution System: A Study on Station Recommendation Using a Genetic Algorithm †

Yang-Chou Juan [1,*], Yi-Chung Chen [1], Wei-Ting Chen [2], Chieh Yang [2], Chia-Tzu Liu [2], Yi-Ci Hou [2] and Yi-Hsuan Tsai [2]

[1] Department of Computer Science and Engineering, National Chung Hsing University, Taichung 402, Taiwan
[2] Department of Industrial Engineering and Management, National Yunlin University of Science and Technology, Yunlin 640, Taiwan
* Correspondence: d112056002@mail.nchu.edu.tw
† Presented at the 31st International Conference on Geoinformatics, Toronto, ON, Canada, 14–16 August 2024.

Abstract: Governments are encouraging public transportation and bicycle-sharing systems to promote sustainable development and reduce greenhouse gas emissions. Despite the expansion of Taipei's YouBike program, many stations frequently run out of bikes or docking spaces, and current redistribution strategies are suboptimal. This study proposes a novel approach to optimize YouBike allocation under resource constraints. We first used K-means clustering to group stations with similar rental profiles, reducing the number of models needed. A random forest model selected key crowd grid factors as input variables for a long short-term memory (LSTM) prediction model to accurately predict demand patterns, including during special events or weather changes. A genetic algorithm then determined optimal station configurations and provided return station recommendations, considering user destinations and station dock ratios, while minimizing manual redistribution. Simulations demonstrated that the proposed system meets user needs, enhances operational efficiency, and significantly reduces manual redistribution costs. Our methods have practical applicability for YouBike managers, indicating that user compliance with recommendations can offset the need for manual redistribution and support the current policy of recommending stations within 600 m of the user's destination.

Keywords: genetic algorithm; long short-term memory (LSTM); YouBike

check for updates

Citation: Juan, Y.-C.; Chen, Y.-C.; Chen, W.-T.; Yang, C.; Liu, C.-T.; Hou, Y.-C.; Tsai, Y.-H. Enhancing Youbike Redistribution System: A Study on Station Recommendation Using a Genetic Algorithm. *Proceedings* **2025**, *110*, 35. https://doi.org/10.3390/proceedings2024110035

Academic Editors: Dongmei Chen, Yuhong He and Songnian Li

Published: 20 February 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Many governments are actively promoting the use of public transportation to further sustainable development and reduce greenhouse gas emissions. Bicycle-sharing systems play an essential role in covering the "last mile" of one's daily commute. Unfortunately, the growing popularity of these systems often results in situations where some stations have no available bikes, while other stations have too many bikes for the available space. Considering the vast number of bicycles and stations in major urban centers, efficiently allocating resources poses a significant logistical challenge.

Researchers have identified a number of factors that hinder the efficiency of the YouBike system in Taipei. The use of fixed routes and real-time bike availability ratios to guide the process of bike redistribution [1,2] is insufficient to overcome a heavy reliance on experienced dispatchers to deal with contingencies, thereby rendering the system susceptible to time lags and decision bias. Efforts to balance the distribution of bikes among multiple stations are hindered by the fact that all decision-making is based entirely on current data without the ability to model behavioral patterns (e.g., weekly commuting), respond to sudden events (e.g., rainstorms), or anticipate future demand, all of which can have a profound impact on bike usage patterns. Figures 1 and 2 illustrate the dramatic increase in crowd flow at the Huashan Cultural and Creative Park during special events and the corresponding YouBike usage at nearby stations.



Figure 1. Crowd flow at Huashan Cultural and Creative Park: (a cultural and creative hub in Taipei, Taiwan): regular daily activity.



Figure 2. Crowd flow in the same location during an event.

In the current study, we applied LSTM models [3–6] to the analysis of crowd-related time series data in order to capture the dynamic features required to make accurate predictions of demand patterns associated with special events or sudden changes in the weather. We also sought to gain the cooperation of users in enhancing system efficiency. One solution involves having users upload their destination to a server tasked with optimizing the allocation of bicycles among stations. Another solution involves the provision of incentives to return bikes to under-supplied stations in order to alleviate the need for manual bike redistribution.

Nonetheless, implementing these solutions presented a number of challenges. Developing LSTM models for every YouBike station would be time-consuming and computationally intractable. Thus, we reduced the number of models by grouping stations according to similarities in usage patterns via K-means clustering [7–9]. The inclusion of crowd-related data in a predictive model greatly increases complexity and variability, both of which can have a detrimental effect on accuracy. Thus, we employed a random forest genetic algorithm to simplify data inputs by identifying the crowd grids with the most pronounced influence on prediction outputs.

The efficacy of the proposed system was assessed in simulations of the public transportation system in Taipei, based on user and station data provided by YouBike. To avoid

skewing the results due to the COVID-19 pandemic, this analysis was performed using data exclusively from 2019.

2. Datasets

The four datasets used in this study included the following: (A) YouBike user rental data for behavior prediction via the LSTM model, (B) YouBike station data to calculate deficiencies in the number of bike stations around destinations, (C) crowd data to enhance LSTM predictions, and (D) weather data to observe the influence of rainfall on YouBike usage behavior.

2.1. YouBike User Database

As shown in Table 1, this dataset comprises 27,956,451 records of Taipei YouBike monthly rentals and user card usage (1 January–31 December 2019), including user card number, rental time, rental station ID, rental station, return time, return station ID, and return station.

Table 1. YouBike rental data.

User Card Number	**Rental Time**	**Rental Station ID**	**Rental Station**
BB99B	6 January 2019 16:34:17	64	National Library
Return Time	**Return Station ID**	**Return Station**	
6 January 2019 16:39:34	327	Chongqing South Rd.	

2.2. YouBike Station Database

As shown in Table 2, this dataset contains data related to 401 Taipei YouBike stations for the same period, including station ID, station name, total parking spaces, available bikes, station area, latitude, longitude, address, available return spaces, and operational status.

Table 2. YouBike station data.

Station ID	**Station Name**	**Total Spaces**	**Available Bikes**
1	MRT City Hall Stn (Exit 3)	180	125
Station Area	**Latitude**	**Longitude**	**Available Return Spaces**
Xinyi Dist.	25.04086	121.5679	55
Address		**Operational Status**	
Zhongxiao East Rd./Songren Rd. (SE)		1	

2.3. Crowd Dataset

This dataset includes crowd flow data for the Taipei area for the same period (2923 data points).

2.4. Weather Dataset

This dataset includes weather data for the Taipei area for the same period.

3. Methods and Results

This study sought to optimize the YouBike resource allocation system by implementing a two-phase process, including offline training and online operations. Figure 3 presents a flow chart of the research process.

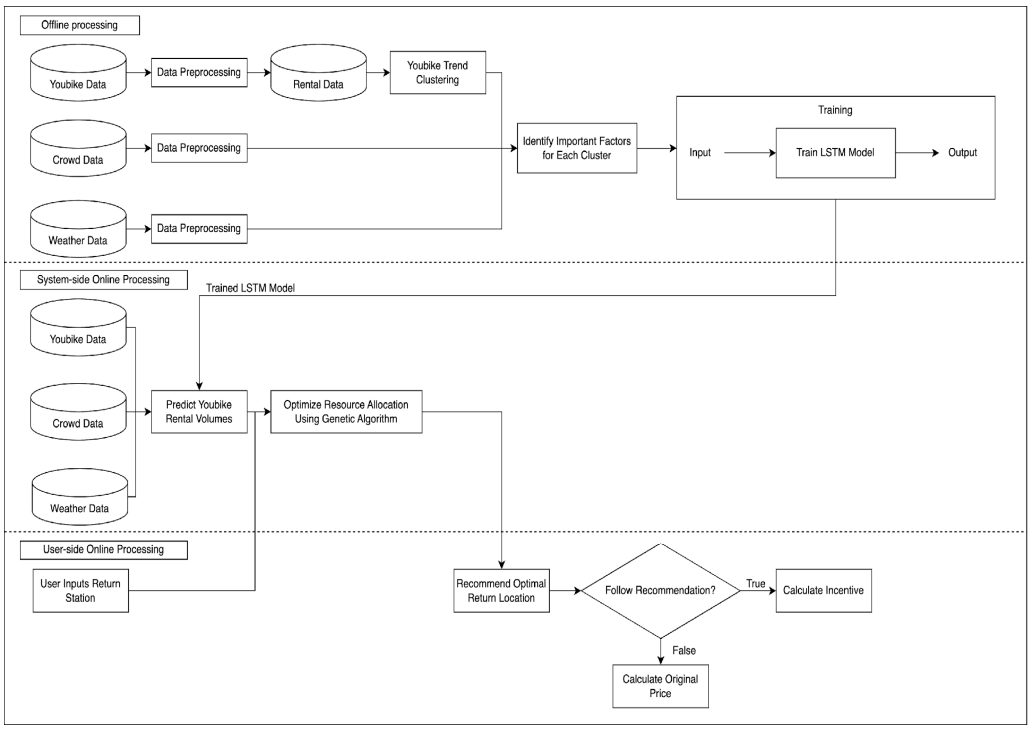

Figure 3. Research flow chart.

Offline processing involved the cleaning of rental and crowd data through the removal of outliers and the filling in of missing values to ensure the quality and reliability of the data. K-means clustering was then used to group the stations according to similarities in bike rentals with the aim of reducing the number of LSTM models and the time required for model training. After Z-score normalization and dimensionality reduction, the silhouette coefficients were used to determine the optimal number of clusters. As shown in Figure 4, any given station could belong to different clusters at different times.

We then used the random forest algorithm to obtain importance scores for the surrounding crowd grid. High-scoring crowd grids were selected as input features for the prediction of bike rental numbers using the LSTM model. For example, for Station 37 (MRT Dongmen Stn, Exit 4), crowd grids E, F, and I generated the highest importance scores (0.166, 0.127, and 0.138) and were therefore employed for predictions (see Figure 5).

Finally, LSTM and Bi-LSTM models were used to predict YouBike rental volumes. The performance of these models in generating predictions for three stations located near a school, MRT station, and hospital (Stations 248, 344, and 63) was evaluated using mean squared error (MSE) and root mean squared error (RMSE) values. The Bi-LSTM model with four hidden layers performed the best, as evidenced by RMSE values ranging from 1 to 3.4 (see Table 3).

Online processing was performed by a genetic algorithm generating bike drop-off recommendations based on the destinations that had been input by users. In addition to the destination data, the algorithm took continuously generated LSTM prediction results of YouBike rental volumes for the next hour as inputs. The process of generating recommendations was a multi-step process, involving chromosome encoding, initial chromosome generation, fitness score calculation, chromosome selection, crossover, and mutation. Of course, this system would not work if the algorithm were unaware of the user's destination; therefore, the proposed system also includes incentives to encourage users' participation in the process of submitting their destinations and complying with the recommendations that they receive. Note that the incentive scheme is meant to reduce the costs of manual redistribution.

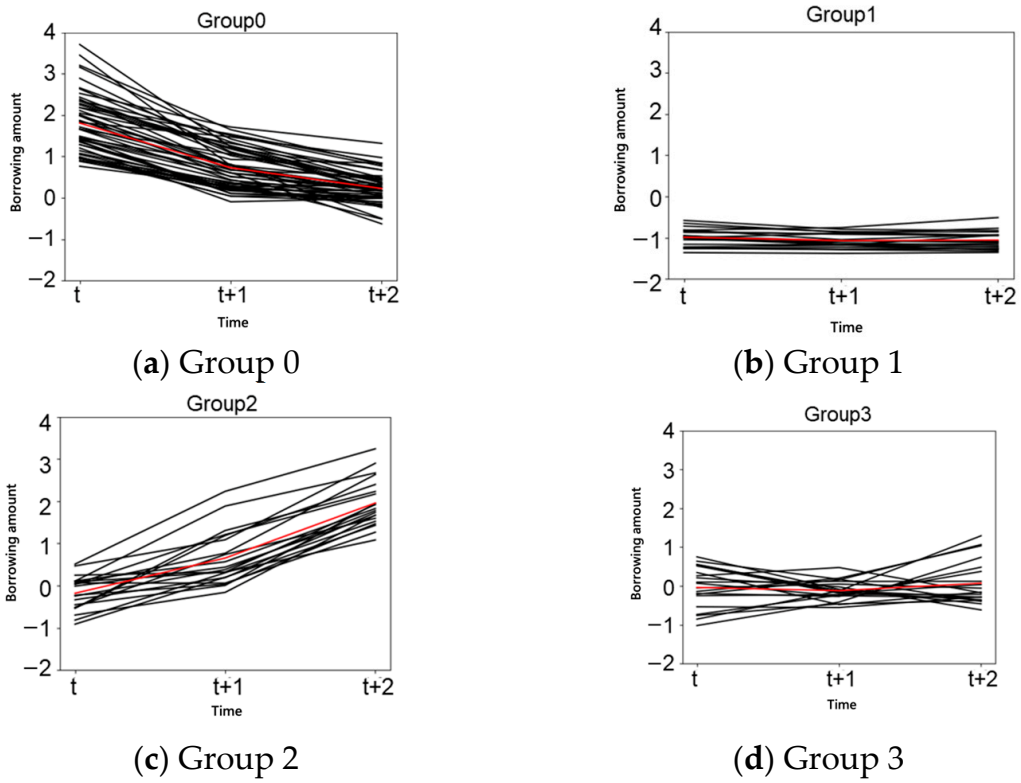

Figure 4. Clustering results.

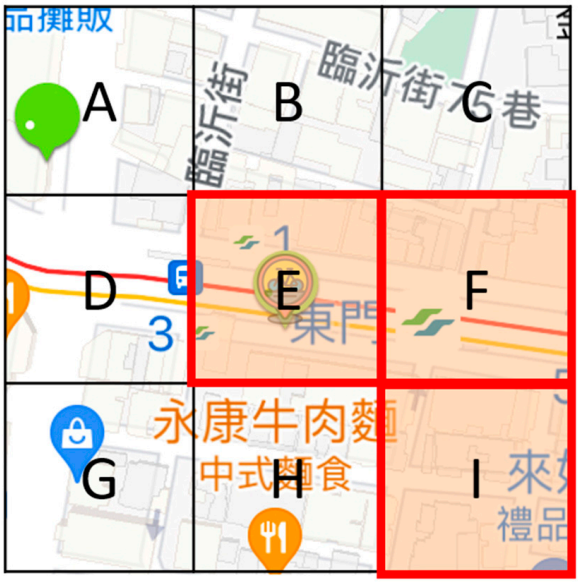

Figure 5. Importance scores of crowd grids around Station 37. Grid E corresponds to MRT Dongmen Stn, Station (Exit 4), while the remaining grids represent surrounding areas.

Table 3. Performance of Bi-LSTM model with four layers.

Station	Station Name	Weekday/Holiday	MSE	RMSE
248	George Industrial School	Weekday	6.3845	2.5268
248	George Industrial School	Holiday	1.1804	1.0864
344	MRT Zhongxiao Fuxing Stn (Exit 5)	Weekday	4.5430	2.1315
344	MRT Zhongxiao Fuxing Stn (Exit 5)	Holiday	2.9807	1.7264
63	Renai Hospital	Weekday	11.2984	3.3613
63	Renai Hospital	Holiday	2.5152	1.5859

The genetic algorithm was optimized by running multiple iterations of a simulation to assess the effects of algorithm parameters, including the number of generations, mutation rate, and crossover rate. Stable results were achieved using 100 generations, and the best learning results were achieved using a crossover rate of 0.7 and a mutation rate of 0.07 (see Figures 6–8).

The performance of the proposed system was assessed quantitatively using simulations based on historical YouBike data in Da'an District, Taipei. As shown in Figure 9, this evaluation involved calculating the number of stations requiring manual adjustments as a function of the user compliance rate (ω%).

The feasibility of the proposed system was also assessed qualitatively using real-world data from 10 users inputting their destinations at noon, with the distance between the recommended return station and the original destination used as a metric. In this hypothetical scenario, the system yielded an adaptation value of 23.1698 when recommending stations within a 600 m radius to ensure user convenience and system efficiency (see Figure 10).

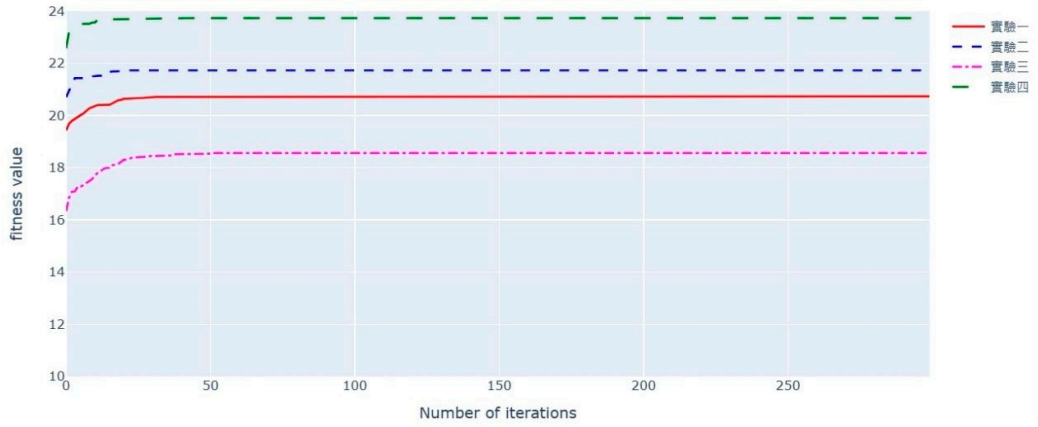

Figure 6. Genetic algorithm: fitness value vs. iterations.

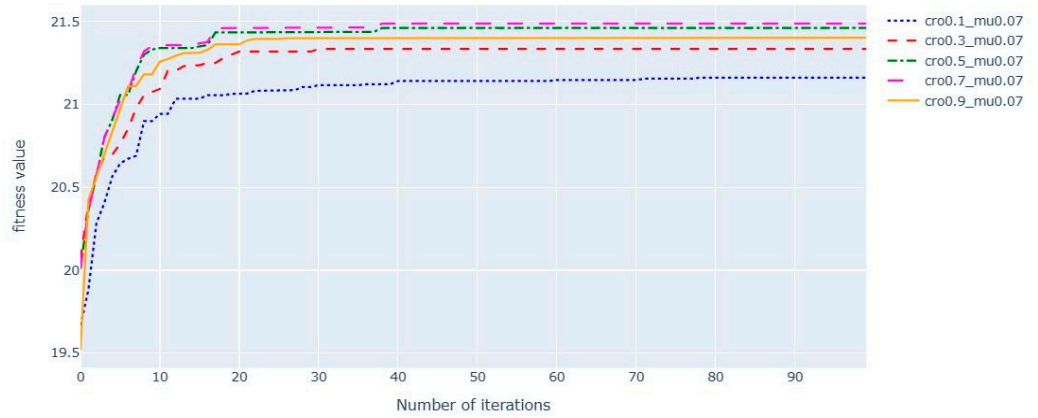

Figure 7. Genetic algorithm: fitness vs. crossover rate.

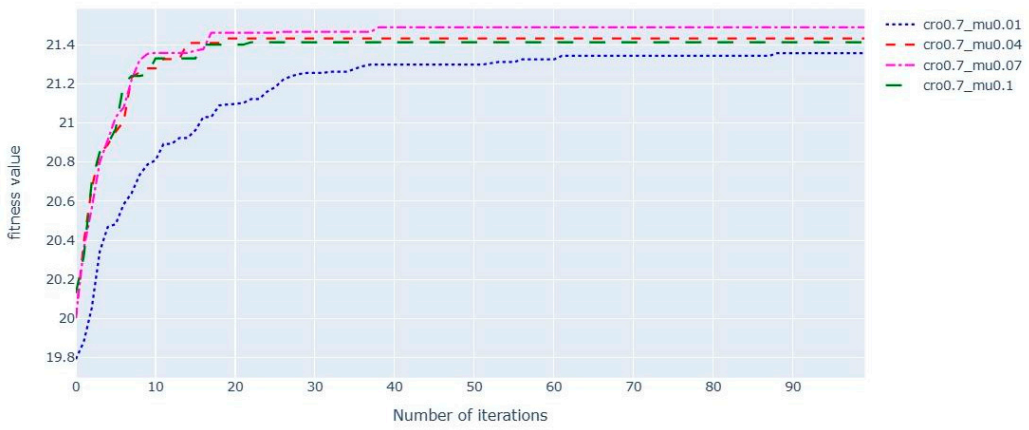

Figure 8. Genetic algorithm: fitness vs. mutation rate (mu).

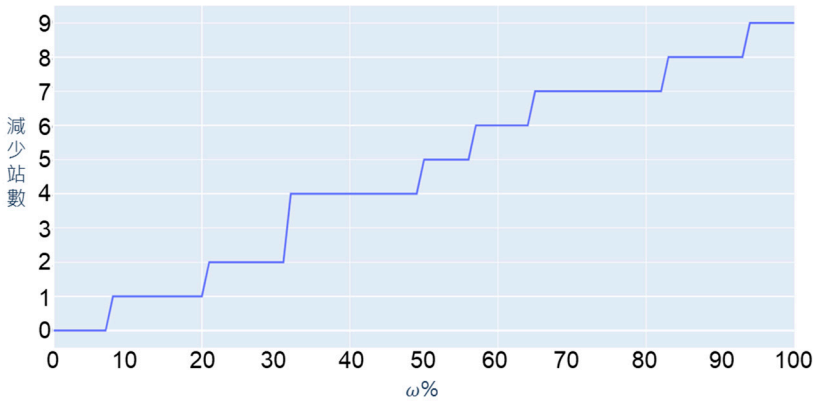

Figure 9. Results of manual adjustment requirements based on user compliance rate.

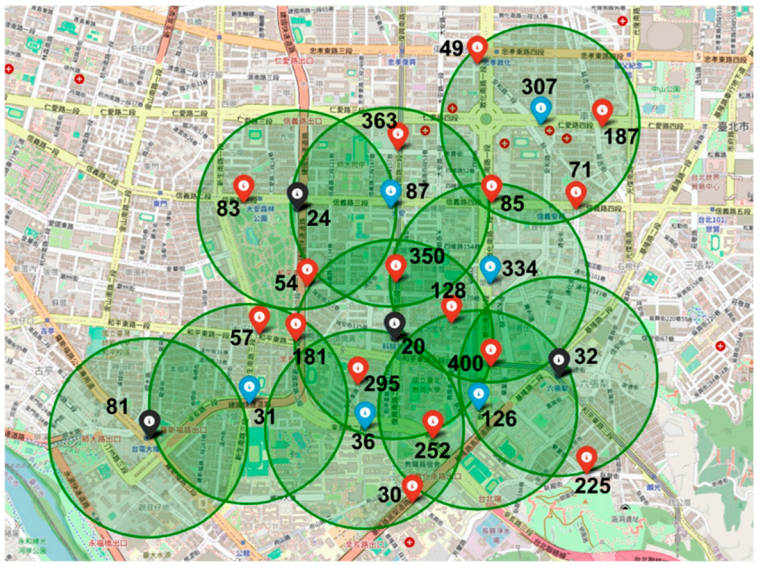

Figure 10. Destinations input by users and corresponding recommendations for the stations at which to return rental bikes.

4. Discussion

This paper proposes a novel approach to the allocation of YouBikes under resource constraints. The K-means algorithm was first used to cluster YouBike stations with similar rental profiles in order to reduce the number of models required. A random forest model

then selected the three crowd grid factors with the most pronounced influence on YouBike rental volumes for use as input variables for the LSTM prediction model. A genetic algorithm then determined the optimal YouBike station configuration and return station recommendations, while taking into account the destinations of users, station bike dock ratios, and the need to minimize the manual redistribution of bikes.

In simulations, the proposed vehicle allocation system was shown to meet user needs and enhance the operational efficiency of the YouBike system, while significantly reducing manual bike redistribution costs. The methods developed in this study have practical applicability in guiding decisions by YouBike managers and staff members.

Future research directions include the assessment of similar methods in other cities in order to verify the generalizability of this scheme. The inclusion of more historical data could help to refine the models even further and in so doing enhance the accuracy of the recommendations. Researchers should also explore other more complex heuristic algorithms and compare their performance with that of the current method.

Author Contributions: Conceptualization, Y.-C.C. and W.-T.C.; methodology, C.Y., C.-T.L., Y.-C.H. and Y.-H.T.; writing—original draft preparation, Y.-C.J.; data curation, Y.-C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study were obtained from Youbike Company and are not publicly available. Access to these data requires a formal request and approval from Youbike Company. Interested readers should contact the corresponding author to initiate the application process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, S.-L. Personal Interview—The Unsung Heroes Behind Convenient YouBike Rentals: 24-Hour Shift Dispatchers. 2020.
2. Yang, X.-H. 2022. Available online: https://news.ltn.com.tw/news/life/paper/1532787 (accessed on 25 November 2024).
3. Zhang, Z.; Chen, P.; Wang, Y.; Yu, G. A hybrid deep learning approach for urban expressway travel time prediction considering spatial-temporal features. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 795–800.
4. He, R.; Liu, Y.; Xiao, Y.; Lu, X.; Zhang, S. Deep spatio-temporal 3D densenet with multiscale ConvLSTM-Resnet network for citywide traffic flow forecasting. *Knowl.-Based Syst.* **2022**, *235*, 109054. [CrossRef]
5. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 865–873. [CrossRef]
6. Ma, X.; Tao, Z.; Wang, Y.; Yu, H.; Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 187–197. [CrossRef]
7. Neethu, B.N.; Jayanthy, S. Greenhouse Monitoring and Controlling using Modified K Means Clustering Algorithm. In Proceedings of the 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 12–14 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 456–462.
8. Ho, C.C.; Ting, C.Y. Time series analysis and forecasting of dengue using open data. In Proceedings of the International Visual Informatics Conference, Bangi, Malaysia, 17–19 November 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 51–63.
9. Wang, Z.; Zhou, Y.; Li, G. Anomaly Detection by Using Streaming K-means and Batch K-means. In Proceedings of the 2020 5th IEEE International Conference on Big Data Analytics (ICBDA), Xiamen, China, 8–11 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 11–17.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.