Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.30642

Research on the Application of Sharing Bicycle Demand Prediction in Urban Planning

Qianyou An[1*], Wenbo Zhao[2]

[1]*Taiyuan Ivy Middle School, Taiyuan, China*
[2]*Shandong Jiaozhou No.1 Middle School, Jiaozhou, China*
**Corresponding Author. Email: aqy043112@gmail.com*

Abstract. This research applies several deep learning models for shared bike demand prediction. By selecting and analyzing a dataset from Kaggle, this research pays attention to identifying factors that have significant impact on bike sharing demands and utilizing models. The methodology includes data description and preprocessing, while the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are used to appraise the model performance, with cross-validation ensuring robustness. Experimental results demonstrate that deep-learning models, especially Transformer and Long-Short Time Memory (LSTM), outperform traditional approaches in both prediction accuracy and adaptability to dynamic rental fluctuations. SHAP value analysis reveals strong nonlinear effects of temperature and time features, while attention weight visualization highlights the Transformer model's ability to detect significant events. Furthermore, deep learning models exhibit greater stability under extreme weather conditions and respond more effectively to sudden demand surges during holidays. These findings contribute to the optimization for bike-sharing operations and development of data-driven urban mobility solutions, which will enhance predictive capabilities in complex urban environments.

Keywords: bike sharing, demand prediction, inter-modal relationship.

1. Introduction

With the development of the sharing economy, urban transportation models are undergoing profound changes. As a significant and fundamental part of urban smart transportation systems, sharing bicycles have emerged as an effective supplement to travel modes. Sharing bicycles not only alleviate urban traffic pressure but also largely reduce carbon emissions and congestion, promoting environmentally friendly travel methods. Therefore, how to precisely predict the demand for sharing bikes has become a key issue in modern urban transportation planning. Apart from being an effective solution to the "last mile" travel problem, sharing bikes are also closely related to the smart city construction. By predicting bike sharing demand, more accurate resource scheduling and management can be implemented, which helps to reduce the phenomena of unbalanced stocking of bicycles and improving the efficiency of operation and commuting for urban employees. Besides, through reasonable demand prediction, data support could promote the collaborative development of public transportation systems.

DOI: 10.54254/2755-2721/2025.30642

Although considerable research has explored the requirement of sharing bike demand prediction, most of them focus on traditional models with linear assumptions relatively. While these models can provide predictions to some extent, they typically neglect to account for the complicated nonlinear factors behind bike-sharing demands, such as extreme weather features, holidays, and particular special events. Generally, traditional linear models are unable to effectively capture these complex relationships multidimensionally, which cause the limitation in the diversity and comprehensiveness of demand prediction [1,2]. In addition, most research in the past decade lacks a dynamic analysis of demand fluctuations.

This research proposes a novel approach to shared bike demand prediction by incorporating nonlinear modeling techniques including multiple factors for different aspects. By incorporating a whole set of factors including extreme weather conditions and temporal characteristics of timing sequence data, this approach significantly improves the ability to predictive demands with higher accuracy. As a result, the deep-learning methods including decision trees, random forests, and gradient boosting can be utilized in this research. The usage of deep learning models such as the Long Short-Term Memory (LSTM) [3] and the Transformer enhances the capability to model temporal dependencies, while the Convolutional Neural Network [4] and LSTM hybrid (CNN-LSTM) [5] further improves the ability to model spatial dependencies in sharing bicycle demands. These innovations provide a more comprehensive and accurate prediction model, offering deeper insights for long-time series data and modeling long-range dependencies. As a result, the relationship between numerous data of shared bike rent count and other non-linear factors can be reflected more directly and perspicuously.

2. Data specification and preprocessing

2.1. Data source and feature description

The Washington DC bike sharing dataset used is sourced and selected from Kaggle, containing details for biking demands across Washington DC bike sharing from 2011 to 2018 [6]. The dataset consists of 11 columns and includes variables—date, temperature of observation, precipitation, wind speed, four distinct weather condition (fog, haze, rain and snow), casual and registered users, the total customers and holiday. The “holiday” and four weather variable are all represented by 1, with 1 indicating “Yes”.

2.2. Feature engineering

To elevate the prediction accuracy of the model, some of feature engineering techniques are applied. First, a time series sliding window approach is used to construct temporal dependencies, applying the model to capture features and trends over different time intervals. Next, categorical variables, such as the day of the week and weather classification, are encoded using one-hot encoding [7] to transform them into numerical representations suitable for deep learning algorithms. Finally, data standardization [8] is utilized to ensure that all numerical features, including temperature, wind speed, and historical bike-sharing rental data, are scaled to a common range. In this research, we use z-score normalization [9], where each feature is transformed using the formula:

$$X' = \frac{X-\mu}{\sigma} \tag{1}$$

DOI: 10.54254/2755-2721/2025.30642

where X' is the calculated value, X is the original value, μ is the mean, and σ is the standard deviation of the feature. This method illustrates that the data owns a mean of 0 and a standard deviation of 1, enhancing model convergence.

3. Methodology

3.1. Baseline model selection

To establish a reference for demand prediction, we implement the following models:

3.1.1. Linear regression (baseline)

Linear regression [10] serves as the simplest predictive model, assuming a linear relationship between input patterns and bike sharing demand.

The model equation is:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{2}$$

where y is the rental prediction demand, β_0 is the original coefficient, x_i are the input features, β_i are the learned coefficients, and ϵ is the error term.

3.1.2. Decision tree and random forest

Decision trees [11] split the dataset based on feature thresholds to make predictions. A decision tree minimizes the impurity at each node by selecting the best feature and threshold.

The impurity is measured using the Gini index [12]:

$$\text{Gini}\left(D\right) = 1 - \sum_{i=1}^{c} {p_i}^2 \tag{3}$$

where D is the dataset at the node, p_i is the proportion of samples belonging to class i, and c is the number of classes.

For regression, the decision tree minimizes the variance:

$$Var\left(D\right) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2 \tag{4}$$

where N is the number of samples, y_i are actual values, and $\overline{y}$ is the mean target value in D.

Random forests [13] improve decision trees by averaging predictions over multiple trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t\left(x\right) \tag{5}$$

where T is the number of trees, $h_t\left(x\right)$ is the prediction from tree t, and $\hat{y}$ is the final prediction.

3.1.3. Support Vector Machines (SVM)

SVM [14] maps input data into higher-dimensional space using different kernel functions (linear, polynomial, RBF) and finds the optimal hyperplane for regression.

DOI: 10.54254/2755-2721/2025.30642

The equation of RBF is as follow:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{6}$$

subject to:

$$y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i$$

$$\langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0 \tag{7}$$

where w and b define the hyperplane, ξ_i and ξ_i^* are slack variables for handling errors, C is a regularization parameter, and ϵ defines the margin of tolerance.

3.1.4. XGBoost

XGBoost [15] applies gradient boosting to build an ensemble of weak learners, optimizing squared loss. It helps capture complex interactions in the data.

The equation of XGBoost is:

$$Obj = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i}\right) + \sum_{t=1}^{T} \Omega(f_t) \tag{8}$$

where $l\left(y_i, \widehat{y_i}\right)$ is the loss function (i.e., squared error for regression), $\Omega(f_t)$ is the regularization term, and T is the number of boosting rounds.

The model is updated using the second-order Taylor expansion:

$$Obj^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{9}$$

where g_i and h_i are the first and second derivatives of the loss function.

3.2. Deep learning model architecture

3.2.1. LSTM (Long Short-Term Memory) networks

LSTM is used for time sequence modeling due to its ability to capture long-term dependencies. The network consists of:

a. Input layers for historical rental data and weather features

b. LSTM layers with hidden units for sequential learning

c. Fully connected layers for prediction output

The update process follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{10}$$

DOI: 10.54254/2755-2721/2025.30642

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{11}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{12}$$

where i_t , f_t , and o_t are the input, forget, and output gates, respectively, x_t is the input vector, h_{t-1} represents the previous hidden state, W_i, W_f, W_o are the weight matrices for input to gate units, U_i, U_f, U_o are the weight matrices for hidden state to gate units, $b_{i,} b_f, b_o$ are the bias terms, and σ is the sigmoid activation function.

3.2.2. CNN-LSTM hybrid model

Convolutional layers extract spatial features from time-series data, which are then processed by LSTM for sequential learning. The CNN extracts high-level patterns while LSTM captures temporal dependencies.

The convolution operation is defined as:

$$F_{i,j} = \sum_m \sum_n X_{i+m,j+n} K_{m,n} \tag{13}$$

where $F_{i,j}$ is the output feature map, X is the input matrix while K is the kernel matrix.

3.2.3. Transformer model

Transformers use self-attention mechanisms to weigh various time steps in the input sequence, enhancing temporal feature extraction.

The Attention mechanism is defined as:

$$\text{Attention}\left(Q, K, V\right) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{14}$$

where Q (queries), K (keys), and V (values) are transformed input representations, and d_k is the dimension of the key vectors. The softmax function normalizes attention weights.

3.3. Comparative experimental design

3.3.1. Unified training and test split

A time-series cross-validation approach is used to split the data while preserving the temporal order.

3.3.2. Evaluation metrics

We assess models using the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \widehat{y_i}\right)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left|y_i - \widehat{y_i}\right| \tag{16}$$

DOI: 10.54254/2755-2721/2025.30642

where n is the number of the sample, y_i is the actual value, $\widehat{y_i}$ is the predicted value, $y_i - \widehat{y_i}$ represents the error, and $|y_i - \widehat{y_i}|$ represents the absolute error.

This methodology ensures a comprehensive evaluation of different models for bike sharing demand prediction.

4. Test result and analysis

4.1. Model performance comparison

4.1.1. Model contrast

To assess the predictive accuracy of using models, we compare their RMSE and MAE. The superior performance of deep learning models over traditional machine learning models is demonstrated.

Table 1. Performance metrics of various models

Model	RMSE(Test)	MAE(Test)	Training Time (CPU)	Training Time (GPU)
Linear Regression	120.5	85.3	2s	N/A
Decision Tree	115.2	80.7	5s	N/A
Random Forest	105.3	75.2	30s	N/A
SVM	110.8	78.5	45s	N/A
XGBoost	98.7	70.1	25s	15s
LSTM	90.2	65.3	120s	45s
Transformer	88.5	63.8	150s	50s

4.1.2. Training time costing comparison

The deep learning models, especially LSTM and Transformer, outperform traditional models, achieving lower RMSE and MAE values. Additionally, GPU acceleration significantly reduces training time, making deep learning approaches more practical for large datasets.

4.1.3. Overfitting analysis

To assess overfitting, we compare training and test errors. Traditional models like decision trees exhibit higher test errors, suggesting overfitting. Conversely, XGBoost and deep learning models demonstrate better generalization, with a relatively smaller error gap between training and test sets.

4.2. Key factors effect analysis

4.2.1. SHAP value visualization

We use SHAP (SHapley Additive exPlanations) [16] to interpret feature importance. Results show that:

a. Temperature has a non-linear impact on bike demand, with demand peaking at moderate temperatures (~20°C) and declining at extreme temperatures.

b. Time of day strongly affects demand, with peaks during commuting hours (7-9 AM, 5-7 PM).

DOI: 10.54254/2755-2721/2025.30642

4.2.2. Attention mechanism in Transformer model

Transformer's attention visualization reveals that:

The model assigns higher weights to weather variables on extreme weather days (rain, snow).

Special events (public holidays) receive strong attention, impacting demand spikes.

4.2.3. Holiday effect on different models

Traditional models (Linear Regression, Decision Tree) fail to capture holiday-induced demand surges. However, deep learning models, especially Transformer, effectively recognize holiday patterns, leading to improved demand predictions.

4.3. Specific scenario validation

4.3.1. Extreme weather stability

To assess robustness, models were tested on extreme weather days (heavy rain, snowfall). Results indicate:

a. Traditional models show a 20-30% RMSE increase in extreme conditions.

b. XGBoost and deep learning models perform more stably, with only a 5-10% RMSE increase, demonstrating their ability to handle unusual demand fluctuations.

4.3.2. Response to holiday demand surges

Comparing model responses during holidays:

a. Transformer correctly captures demand spikes, leading to 9.2% lower prediction error compared to other models.

b. Traditional models underestimate demand surges, while deep learning-based methods adapt to irregular demand fluctuations better.

5. Conclusion and future outlook

5.1. Key research findings

This research analysed various deep learning models for shared bike demand prediction using historical rental data from Washington, D.C. Our findings highlight that deep-learning models, particularly LSTM and Transformer, outperform traditional models in capturing complex temporal dependencies and adapting to special situations, such as extreme weather conditions and holiday demand surges. While XGBoost and Random Forest offer a balance between accuracy and computational efficiency, they struggle to generalize under atypical conditions. The SHAP value analysis confirmed the strong nonlinear effects of temperature and time features on the amount of bike rentals, and attention weight visualization in the Transformer model demonstrated its ability to focus on critical time periods, improving its predictive performance.

5.2. Recommendations for smart urban transportation systems

The insights from this research have direct implications for smart urban transportation planning. City planners and bike-sharing operators can leverage deep learning-based predictive models to dynamically allocate bicycles, ensuring availability in high-demand areas, especially during peak

DOI: 10.54254/2755-2721/2025.30642

hours and holiday seasons. Additionally, integrating these predictive models with real-time data streams—such as GPS tracking, live weather feeds, and commuter flow—can further enhance the accuracy of demand predictions and improve system efficiency. Automated rebalancing strategies can be developed based on model predictions, reducing operational inefficiencies and enhancing users' satisfaction.

5.3. Future improvements and research directions

Despite the promising results, several improvements can be explored in future work:

a. Real-Time Data Integration – The current models rely on historical data, but incorporating real-time traffic, weather, and event data can significantly improve forecasting accuracy. Streaming frameworks like Apache Kafka can be used for handling live data ingestion and processing.

b. Dynamic Mechanism Framework – A more adaptive system can be developed using reinforcement learning-based dynamic mechanisms, where the model continuously updates based on evolving rental patterns, optimizing for demand fluctuations.

c. Graph Neural Networks (GNN) – Given that bike-sharing systems function in a spatially interconnected network, GNNs can be explored to model complex station-to-station dependencies. By considering station locations and ridership flows as a graph structure, GNNs can enhance prediction accuracy in highly dynamic urban environments.

d. Multi-Modal Fusion – Future models should leverage multi-modal data fusion, incorporating satellite imagery, social media trends, and public transit usage alongside traditional structured data. This can provide a comprehensive contextual understanding of rental demand patterns, particularly in response to large-scale events or policy changes.

By integrating these advancements, future bike-sharing prediction models can become more adaptive, real-time, and context-aware, supporting the development of smarter, data-driven urban mobility solutions.

References

[1] C. Xu, J. Ji, and P. Liu, “The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets, ” Transportation research part C: emerging technologies, vol. 95, pp. 47-60, 2018.
[2] K. Jin, W. Wang, S. Li, P. Liu, and H. Sun, “Dockless shared-bike demand prediction with temporal convolutional networks, ” in CICTP 2020, pp. 2851-2863, 2020.
[3] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232, 2016.
[4] Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019, 2021.
[5] J.-H. Cho, S. W. Ham, and D.-K. Kim, “Enhancing the accuracy of peak hourly demand in bike-sharing systems using a graph convolutional network with public transit usage data, ” Transportation Research Record, vol. 2675, no. 10, pp. 554-565, 2021.
[6] Kaggle datasets. Available at https: //www.kaggle.com/datasets/gauravduttakiit/bike-sharing.
[7] Rodríguez, P., Bautista, M. A., Gonzalez, J., and Escalera, S. Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing, 75, 21-31, 2018.
[8] Shanker, M., Hu, M. Y., and Hung, M. S. Effect of data standardization on neural network training. Omega, 24(4), 385-397, 1996.
[9] Fei, N., Gao, Y., Lu, Z., and Xiang, T. Z-score normalization, hubness, and few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 142-151, 2021.
[10] C. Rudloff and B. Lackner, “Modelling demand for bike-sharing systems: neighbouring stations as source for demand and reason for structural breaks, ” Transportation Research Record, vol. 2430, no. 1, pp. 1-11, 2014.
[11] De Ville, B. Decision trees: Wiley Interdisciplinary Reviews: Computational Statistics, 5(6), 448-455, 2013.

DOI: 10.54254/2755-2721/2025.30642

[12] Daniya, T., Geetha, M., and Kumar, K. S. Classification and regression trees with Gini index. Advances in Mathematics: Scientific Journal, 9(10), 8237-8247, 2020.
[13] S. Guidon, D. J. Reck, and K. Axhausen, “Expanding a (n)(electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests, ” Journal of Transport Geography, vol. 84, p. 102692, 2020.
[14] Chandra, M. A., and Bedi, S. S. Survey on SVM and their application in image classification. International Journal of Information Technology, 13(5), 1-11, 2021.
[15] S. Feng, H. Chen, C. Du, J. Li, and N. Jing, “A hierarchical demand prediction method with station clustering for bike sharing system, ” in 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE, pp. 829-836, 2018.
[16] Van den Broeck, G., Lykov, A., Schleich, M., and Suciu, D. On the tractability of SHAP explanations. Journal of Artificial Intelligence Research, 74, 851-886, 2022.