International Journal of Economics, Finance and Management Sciences
2024, Vol. 12, No. 5, pp. 258-266
https://doi.org/10.11648/j.ijefm.20241205.13

Research Article

A Soft Voting Ensemble Model for Hotel Revenue Prediction

Yuxin Jiang, Chengjie Ni*, Mingjing Chen

School of Information Technology and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China

Abstract

In recent years, the hotel industry has faced unprecedented opportunities and challenges due to the increasing demand for travel and business trips. This growth not only presents significant opportunities but also brings challenges to resource management and price setting. Accurate hotel revenue prediction is crucial for the hotel industry as it influences pricing strategies and resource allocation. However, traditional hotel revenue prediction models fail to capture the diversity and complexity of hotel revenue data, resulting in inefficient and inaccurate predictions. Then, with the development of the ensemble learning, its application to hotel revenue prediction has emerged as an influential research direction. This study proposes a soft voting ensemble model for hotel revenue prediction, which includes six base models: Convolutional Neural Network, K-nearest Neighbors, Linear Regression, Long Short-term Memory, Multi-layer Perceptron, and Recurrent Neural Network. Firstly, the hyper-parameters of the base models are optimized with Bayesian optimization. Subsequently, a soft voting ensemble method is used to aggregate the predictions of each base model. Finally, experimental results on the hotel revenue dataset demonstrate that the soft voting ensemble model outperforms base models across six key performance metrics, providing hotel managers with more accurate revenue prediction tools to aid in scientific management decisions and resource allocation strategies. This study confirms the effectiveness of the soft voting ensemble model in enhancing the accuracy of hotel revenue forecasts, demonstrating its significant potential for application in strategic planning within the modern hotel industry.

Keywords

Soft Voting, Ensemble Model, Hotel Revenue, Prediction

1. Introduction

With the improvement in the pandemic situation, there has been a notable increase in tourism and business travel in recent years. Hotels have gained popularity due to the convenience and comfort they provide for tourists, business trips, special events, and temporary accommodations. People choose to stay in hotels not only for the ease and comfort during their travels but also because hotels can cater to their diverse accommodation needs in various situations.

Accurate hotel revenue prediction is crucial for the hospitality industry, as it influences decisions on pricing strategies and resource allocation. The dynamic nature of tourism and the hospitality industry, shaped by factors such as seasonality, local events, and economic conditions, poses challenges for traditional statistical models to capture the complex and non-linear patterns in revenue data. However, the emergence of machine learning and ensemble learning methods offers

*Corresponding author: 1374706274@zufe.edu.cn (Chengjie Ni)

Received: 9 August 2024; **Accepted:** 9 September 2024; **Published:** 11 September 2024

promising solutions to enhance prediction accuracy.

In recent years, ensemble learning methods have garnered significant attention, particularly through soft voting ensemble methods, due to their ability to combine the strengths of multiple predictive models to enhance forecasting performance. Soft voting ensemble methods represent a straightforward yet effective ensemble learning strategy. By aggregating the predictive outcomes of several base models through the soft voting ensemble methods, they yield a final prediction. Compared to individual models, soft voting ensemble methods more effectively capture the diversity and complexity inherent in the data, thereby improving prediction accuracy.

This study proposes an ensemble multimodal fusion method based on the soft voting ensemble methods for accurate prediction of hotel revenue. The process begins with preprocessing raw data, which includes filling in the averages of the two adjacent days for anomalies and null data, and normalizing the data using min-max scaling. Subsequently, six base models - Convolutional Neural Network (CNN), K-Nearest Neighbors (KNN), Linear Regression (LR), Long Short-term Memory (LSTM), Multi-layer Perceptron (MLP), and Recurrent Neural Network (RNN) - are trained and used for prediction, with Bayesian optimization employed to identify the optimal hyperparameters. These base models are then integrated using the soft voting ensemble method, and the performance of the models is evaluated based on several metrics. The evaluation demonstrates that the ensemble model outperforms individual machine learning models in terms of prediction accuracy.

The remainder of this paper is organized as follows. Section 2 reviews related work in the fields of hotel revenue prediction and ensemble learning. Section 3 describes the data preprocessing steps and the methodology for constructing the soft voting ensemble model. Section 4 presents the experimental results and analysis. Finally, Section 5 concludes the study and discusses future research directions.

2. Related Work

2.1. Deep Learning for Hotel Revenue

Traditional predictive methods, which frequently depend on simple statistical models, often struggle with complex data and fail to capture nonlinear relationships and hidden patterns within the data. In contrast, deep learning models, owing to their advanced feature extraction and sequential data processing capabilities, are extensively employed in various time series prediction tasks.

For example, Lim et al. [1] compared the predictive capabilities of the Autoregressive Integrated Moving Average (ARIMA) model with Artificial Neural Networks (ANN) in forecasting the prices of apartments in Singapore, and found that the ANN model outperformed the ARIMA model in predictive accuracy. Zhang and Niu [2] combined LSTM networks with CNN to forecast the hotel demand, highlighting the substantial potential of deep learning technology in time series prediction. Noomesh and Baby [3] employed a hybrid CNN-LSTM model to predict tourism demand, which achieved a comprehensive understanding of data patterns and enhanced the precision of the predictions. Moreover, deep learning frameworks enable the integration of multiple data sources, which can significantly enhance the predictive power [4].

This study further validates the efficacy of hybrid deep learning models in handling complex and nonlinear data, offering new perspectives and tools for future hotel revenue management and optimization strategies. The study also find that the deep learning has the capacity to process and learn from a vast array of diverse data [5], making it more comprehensive in understanding the factors that influence hotel revenue.

2.2. Ensemble Method

Ensemble method has emerged as a prominent research direction in the field of machine learning in recent years. It enhances the accuracy and robustness of predictions by amalgamating multiple models. This approach has been widely applied across various sectors such as chemistry, healthcare, finance, and transportation [6-9], demonstrating the significant utility in reducing prediction errors by integrating the strengths of diverse models [10-13].

A common ensemble technique, called the soft voting ensemble method, has often been used to aggregate the predictions of several classifiers to produce a final outcome through a voting mechanism. Compared to other ensemble methods, the soft voting ensemble method is more flexible and accurate. It can balance noise among different models, thereby enhancing the generalization ability of the models, effectively reducing the prediction errors of individual models and enhancing the overall prediction accuracy.

Recent studies have validated the effectiveness of the soft voting ensemble method in numerous fields including computer science, legal technology, and finance. For instance, Mim et al. [14] utilized soft voting ensemble method to detect credit card fraud, demonstrating its effectiveness in handling the imbalanced data with superior performance compared to individual models. Guan et al. [15] integrated predictions from multiple machine learning models (specifically Random Forest, Extra Trees, and CatBoost) using the soft voting ensemble method, and improved the accuracy of predicting key paths in labor dispute resolution. Gao et al. [16] demonstrated that soft voting ensemble method can leverage the individual strengths of different models to improve the accuracy of predicting credit risk for small and medium-sized enterprises.

In summary, the combination of deep learning and ensemble learning methods provides a strong support for hotel revenue prediction. Deep learning models offer the capability to learn from complex and diverse data, while ensemble methods ensure robustness and accuracy in predictions. These advancements represent a significant step forward in developing reliable and effective hotel revenue prediction models.

3. Methodology

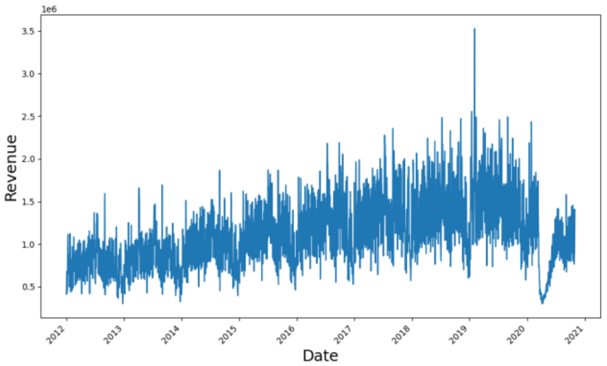
3.1. Dataset Description

The data utilized in this experiment consists of daily revenue records from mid to high end hotels located in Atlanta, Boston, Chicago, Dallas, Los Angeles, New York, Orlando, and Washington, spanning the years 2013 to 2020. The dataset comprises a total of 23,255 entries. Among the eight available nodes, this study focuses exclusively on the "Revenue" node for analysis. The nature of the data is well-suited for time series prediction. The data source can be accessed at https://paperswithcode.com/dataset/hotel-sales.

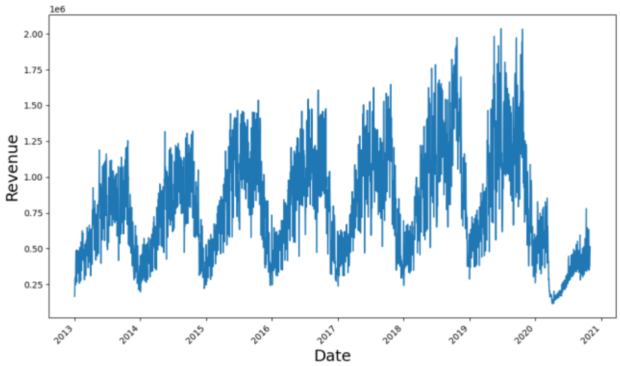
3.2. Data Exploration and Preprocessing

3.2.1. Data Exploration

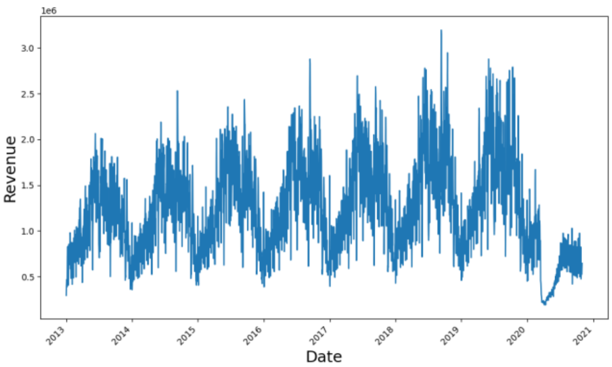
First, missing values in the dataset were addressed by imputing them with the average values of the adjacent two days. Subsequently, the hotel revenue data from the eight regions were visualized, as shown in Figure 1.



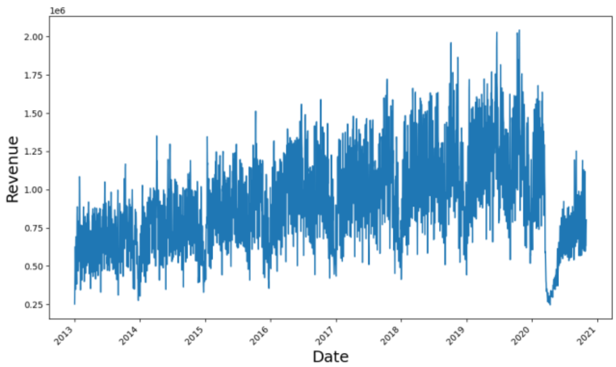
a) Revenue over Date for Atlanta



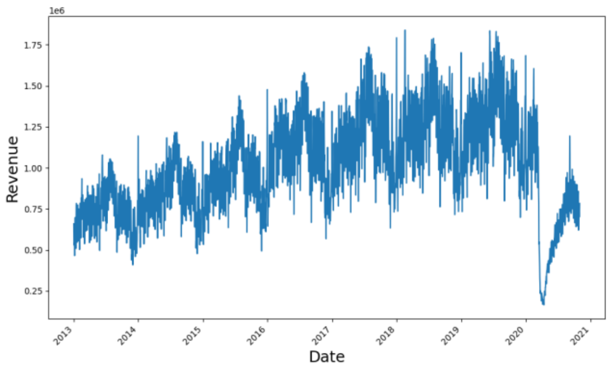
b) Revenue over Date for Boston



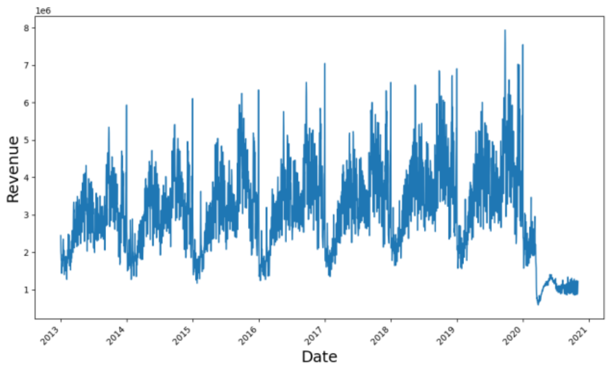
c) Revenue over Date for Chicago



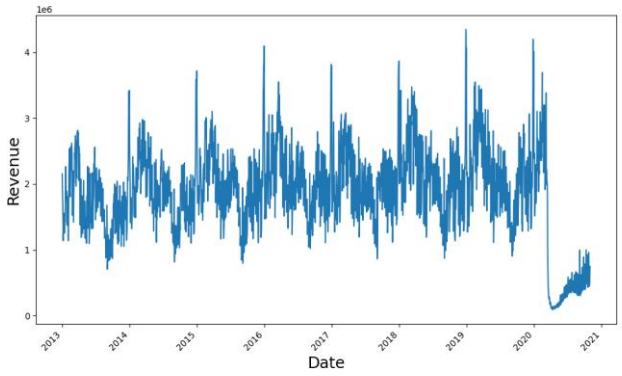
d) Revenue over Date for Dallas



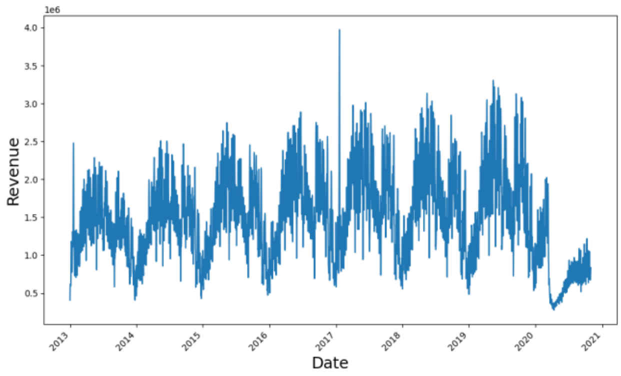
e) Revenue over Date for LosAngeles



f) Revenue over Date for NewYork



g) Revenue over Date for Oriando



h) Revenue over Date for Washington

Figure 1*. Revenue situation of hotels in eight regions.*

Figure 1 reveals a distinct cyclical pattern in hotel revenue data. Generally, revenue is lower at the beginning and end of the year, while mid-year marks the peak revenue period. There is a noticeable decline in hotel revenue in 2020, likely attributable to the impact of the COVID-19 pandemic. Additionally, the data visualization identified some anomalies. To accurately address these anomalies, considering the impact of specific U.S. public holidays and data dimensions, the Moving Window Method was employed for detection.

3.2.2. Data Preprocessing

In this study, the Moving Window Method was utilized to detect anomalies within the dataset, and any detected anomalies were imputed with the average values of the adjacent two days.

The Moving Window Method involves applying a fixed-size window to time series data, calculating statistical measures within each window, and using these measures to determine if a data point is an anomaly.

The mean and standard deviation were calculated for each window. If a data point's deviation from the mean exceeds a specified threshold (k times the standard deviation), it was classified as an anomaly. In this study, the threshold k was set to 3 for non-holidays and 6 for holidays, with a window size of 10, meaning it covers the data points from the five days before and after the point of interest. The window was continuously moved until all data points were evaluated.

The mathematical representation is provided in Equations (1) and (2) as follows:

Mean and Standard Deviation Calculation:

For each window w_i:

$$\mu_i = \frac{1}{n}\sum_{x_i \in w_i} x_i \tag{1}$$

$$\sigma_i = \sqrt{\frac{1}{n}\sum_{x_i \in w_i}(x_i - \mu_i)^2} \tag{2}$$

Anomaly Detection:

For each data point x_i in the window w_i: If $|x_i - u_i| > k_i * \sigma_i$ then x_i is an anomaly, where x_i represents the i-th data point in the window w_i; n represents the number of data points in window w_i; w_i represents the window centered at x_i, including x_{i-5}, x_{i-4}, …, x_i, …, x_{i+4}, x_{i+5}; μ_i and σ_i are the mean and standard deviation of the data points within the window w_i; k_i is the threshold factor, with k_i = 3 for non-holidays and k_i = 6 for holidays.

After detecting anomalies, they were replaced with the average revenue data from the adjacent two days. Subsequently, the data were normalized.

Data normalization plays a crucial role in deep learning and ensemble learning, as it can accelerate convergence, stabilize gradient descent, enhance model performance, and prevent numerical instability. By scaling the data to a smaller range, normalization ensures that each feature contributes equally to the model, facilitating better capture of relationships among features. In ensemble learning, normalization improves model compatibility, reduces bias and variance, and enhances overall model performance and generalization capability. This study employs the Min-Max normalization, as shown in Equation (3), to scale the data to a range between 0 and 1:

$$X\left(normalized\right) = \frac{x - \min\left(x\right)}{\max\left(x\right) - \min\left(x\right)} \tag{3}$$

To comprehensively evaluate the model, the hotel revenue data, which has been normalized using Min-Max normalization, is divided into three segments: 64% was allocated to the training set for model training, 16% to the validation set for hyperparameter tuning and model selection, and the remaining 20% to the test set for the final assessment of model performance.

3.3. Ensemble Model

The workflow of this section is shown in Figure 2.

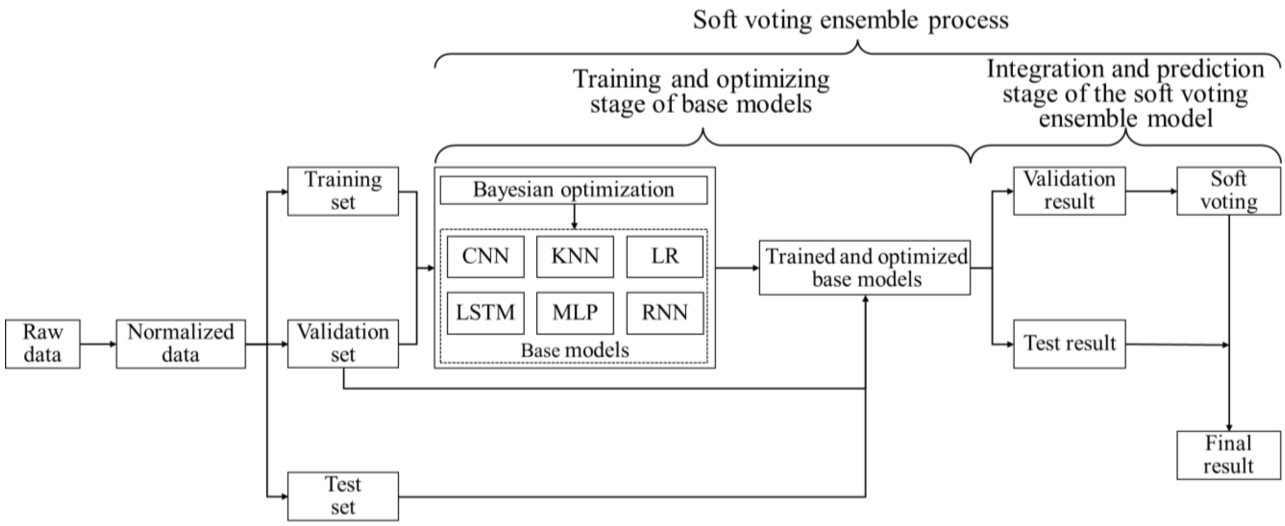


Figure 2. *Schematic overview of the soft voting ensemble model.*

(1) Utilization of Bayesian Optimization

In the context of hotel revenue data, which exhibits significant time series characteristics, traditional grid search and random search methods are inefficient to deal with high-dimensional hyperparameter spaces. Bayesian optimization is a global optimization algorithm based on probabilistic models, widely used for hyperparameter optimization of base models. In this study, a Gaussian process is selected as the prior probability distribution. The procedural framework of Bayesian optimization for hyperparameter tuning is structured into three distinct phases: selecting the optimal evaluation point based on the optimization objective, incorporating the objective function into the observed data, and updating the posterior distribution and the expected improvement function of the objective function. Bayesian optimization efficiently performs global searches in the hyperparameter space, reducing computational costs and quickly converging to the optimal or near-optimal solutions. This approach significantly enhances the efficiency and effectiveness of model training, especially for tasks involving time series data like hotel revenue.

(2) Training and Optimization of Base Models

Considering the characteristics of hotel revenue data, this study selects six base models: CNN, KNN, LR, LSTM, MLP, and RNN. These models are trained and optimized using the Bayesian optimization algorithm to identify the optimal hyperparameters.

(3) Architecture of the Soft Voting Ensemble Model

The architecture of the soft voting ensemble model aims to combine the strengths of each base model to improve overall prediction performance. Specifically, the soft voting ensemble model is composed of two layers: In the first layer, each base model independently predicts the data, and the optimal hyperparameters are identified using Bayesian optimization to generate prediction results. In the second layer, the soft voting ensemble mechanism integrates the predictions from the first layer's base models by averaging their outputs using a soft voting ensemble method, thereby producing the final prediction. This architecture allows the soft voting ensemble model to effectively integrate the predictive capabilities of the base models, capturing the multiple patterns and temporal dependencies in hotel revenue data, and significantly enhancing prediction accuracy and robustness.

4. Experiments

To comprehensively evaluate the models, this section introduces six statistical metrics used to assess model performance. These metrics were employed to analyze the performance of each model in predicting hotel revenue, and illustrated below. All models and methods were implemented using the Python programming language.

4.1. Evaluation Metrics

Before providing the definitions of evaluation metrics, it is essential to clarify the notations used in the equations. In these equations, y_i represents the true value of the i-th sample point, y_i represents the predicted value of the i-th sample point, n represents the number of samples, and $\bar{y}$ represents the mean of the actual values.

(1) Mean Absolute Error (MAE)

MAE represents the average of the absolute differences between the predicted and actual values. It is calculated in Equation (4):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i\right| \tag{4}$$

MAE provides an intuitive measure of the average deviation between predicted and actual values. A lower MAE indicates higher prediction accuracy. One advantage of MAE is that it assigns equal weight to each error term, making it easy to understand and interpret.

(2) Mean Absolute Percentage Error (MAPE)

MAPE is the average of the absolute percentage errors, calculated in Equation (5):

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i}{y_i}\right| \tag{5}$$

MAPE standardizes the error as a percentage, facilitating comparison across different scales. A lower MAPE indicates a smaller average relative error.

(3) Root Mean Squared Error (RMSE)

RMSE is the square root of the average of the squared differences between predicted and actual values, calculated in Equation (6):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2} \tag{6}$$

RMSE amplifies larger errors due to the squaring term, making it more sensitive to outliers. A lower RMSE indicates higher prediction accuracy. This metric effectively reflects the overall prediction performance, especially for models with significant errors.

(4) R-squared (R^2)

R^2 is a statistical measure of the proportion of variance in the dependent variable that is predictable from the independent variables, calculated in Equation (7):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{7}$$

The value of R^2 ranges from 0 to 1, with values closer to 1 indicating a better fit. R^2 provides an intuitive measure of how well the model explains the data, and it is widely used in regression analysis.

(5) Mean Absolute Scaled Error (MASE)

MASE is a scale-independent metric for comparing forecast errors, calculated in Equation (8):

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i\right|}{\frac{1}{n-1}\sum_{i=2}^{n}\left|y_i - y_{i-1}\right|} \tag{8}$$

A lower MASE indicates higher prediction accuracy. The advantage of MASE is its standardization, allowing for more intuitive and fair comparisons across different datasets and models.

(6) Symmetric Mean Absolute Percentage Error (sMAPE)

sMAPE is an improved version of MAPE that avoids the issues of zero values and extreme values, calculated in Equation (9):

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i\right|}{\left(\left|y_i\right| + \left|y_i\right|\right)/2} * 100\% \tag{9}$$

sMAPE symmetrically treats overestimates and underestimates, avoiding the potential pitfalls of MAPE. A lower sMAPE indicates a smaller average proportional error.

4.2. Experimental Results Analysis

Based on the above evaluation metrics, the performance of the six base models - CNN, KNN, LR, LSTM, MLP, and RNN - along with the ensemble learning model based on the soft voting ensemble method, can be observed in Table 1 and Figure 3.

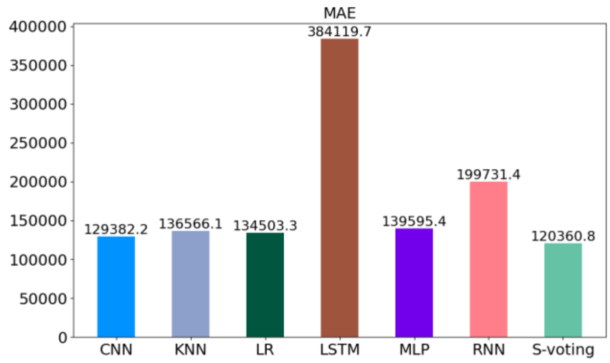
The results, as illustrated by Table 1 and Figure 3, substantiate the effectiveness of the soft voting ensemble model. It is clear that the soft voting ensemble model, which combines the base models CNN, KNN, LR, LSTM, MLP, and RNN through a soft voting ensemble method, exhibits significant performance improvements. The ensemble model outperforms each individual base model across all metrics, showcasing its superior predictive capabilities for the hotel revenue data used in this study. Consequently, it can be concluded that the soft voting ensemble method, by leveraging the strengths of multiple models, effectively enhances prediction accuracy and model robustness. This provides hotel managers with more precise revenue forecasting tools, facilitating more informed business decisions and resource allocation.

5. Conclusion and Future Work

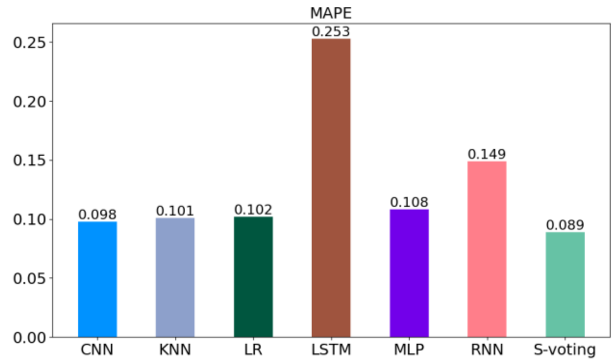
This study proposes a soft voting ensemble learning model for hotel revenue prediction. The model integrates six base models - CNN, KNN, LR, LSTM, MLP, and RNN - optimized through Bayesian optimization, using a soft voting ensemble method to enhance prediction accuracy. The performance of the ensemble model was evaluated using six key metrics: MAE, MAPE, RMSE, R^2, MASE, and sMAPE. The experimental results indicate that the soft voting ensemble model outperforms individual base models across multiple evaluation metrics, demonstrating significant advantages in hotel revenue prediction.

This study also validates the effectiveness of the deep learning models and ensemble learning model in handling complex, nonlinear time series data, highlighting its potential and practical value in hotel revenue prediction. These findings provide hotel managers with more accurate revenue forecasting tools, aiding in the formulation of more informed business strategies and resource allocation plans.

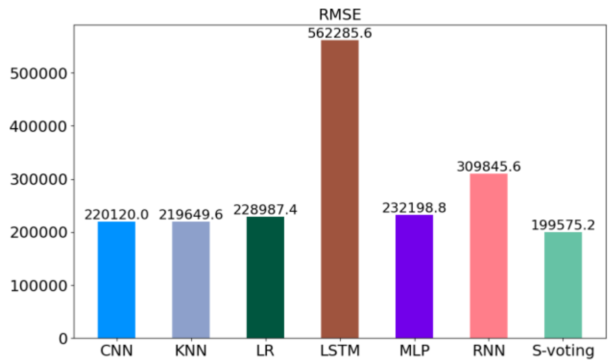
For future research, we can further optimize the parameters of the ensemble model, explore additional combinations of base models, and incorporate more data features to improve predictive performance. Additionally, this method can be applied to other domains, such as tourism and retail, to verify its generalizability and applicability.



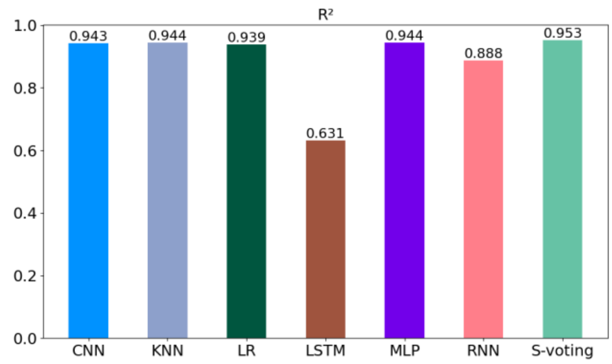
a) Evaluation results using MAE



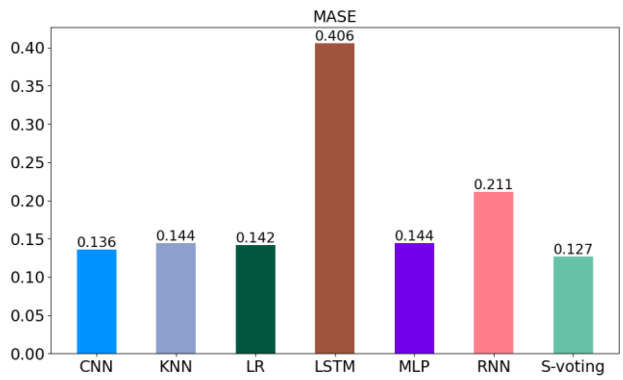
b) Evaluation results using MAPE



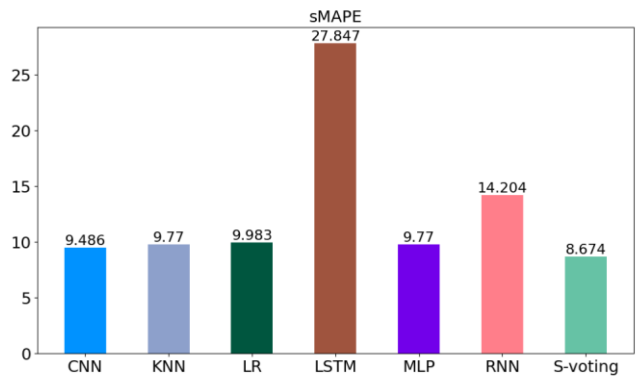
c) Evaluation results using RMSE



d) Evaluation results using R^2



e) Evaluation results using MASE



f) Evaluation results using sMAPE

Figure 3*. Evaluation results.*

Table 1*. The comparative results among various models.*

Metrics	MAE	MAPE	RMSE	R^2	MASE	sMAPE
CNN	129382.188	0.098	220119.972	0.943	0.136	9.486
KNN	136566.115	0.101	219649.635	0.944	0.144	9.770
LR	134503.255	0.102	228987.373	0.939	0.142	9.983
LSTM	384119.705	0.253	562285.614	0.631	0.406	27.847
MLP	139595.383	0.108	232198.827	0.944	0.144	9.770
RNN	199731.356	0.149	309845.634	0.888	0.211	14.204
S-voting	120360.829	0.089	199575.239	0.953	0.127	8.674

Abbreviations

CNN	Convolutional Neural Network
KNN	K-nearest Neighbors
LR	Linear Regression
LSTM	Long Short-term Memory
MLP	Multi-layer Perceptron
RNN	Recurrent Neural Network
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Networks
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Squared Error
R^2	R-squared
MASE	Mean Absolute Scaled Error
sMAPE	Symmetric Mean Absolute Percentage Error

Author Contributions

Yuxin Jiang: Writing - original draft, Methodology
Chengjie Ni: Software, Data curation
Mingjing Chen: Writing - review & editing, Conceptualization

Data Availability Statement

The data that support the findings of this study can be found at: https://paperswithcode.com/dataset/hotel-sales.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Lim, W. T., Wang, L., Wang, Y., & Chang, Q. (2016). Housing price prediction using neural networks. *In Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, Changsha, China, August 13-15, 2016, pp. 518-522. https://doi.org/10.1109/FSKD.2016.7603227

[2] Zhang, D., & Niu, B. Z. (2024). Leveraging online reviews for hotel demand forecasting: A deep learning approach. *Information Processing and Management*, 61(1), 103527. https://doi.org/10.1016/J.IPM.2023.103527

[3] Noomesh, D., & Baby, G. (2023). Forecasting resort hotel tourism demand using deep learning techniques -A systematic literature review. *Heliyon*, 9(7), e18385. https://doi.org/10.1016/j.heliyon.2023.e18385

[4] Jan S. T., & Ahsan C. M. (2020). Deep learning for the internet of things: Potential benefits and use-cases. *Digital Communications and Networks*, 7(4), 526-542. https://doi.org/10.1016/j.dcan.2020.12.002

[5] Bai, J. Y. (2023). Deep learning-based intraoperative video analysis for supporting surgery. *Concurrency and Computation: Practice and Experience*, 35(26), e7837 https://doi.org/10.1002/cpe.7837

[6] Sun, J. L., & Shi, Y. B. (2024). Quantitative analysis of polycyclic aromatic hydrocarbons based on ensemble learning. *Computer Informatization and Mechanical System*, 7(2), 58-60. https://doi.org/10.12250/JPCIAMS2024090313

[7] Zhang, H. Q., Liu, S. H., Li, R., Yu, J. W., Ye, D. X. (2024). MIBPred: Ensemble learning-based metal ion-binding protein classifier. *ACS Omega*, 9(7), 8439-8447. https://doi.org/10.1021/ACSOMEGA.3C09587

[8] Paldino, G. M., Lebichot, B., Le Borgne, Y., Siblini, W., Oblé, F., et al. (2022). The role of diversity and ensemble learning in credit card fraud detection. *Advances in data analysis and classification*, 18(1), 21-25. https://doi.org/10.1007/S11634-022-00515-5

[9] Yan, H., Fu, L. Y., Qi, Y., Yu, D.-J., & Ye, Q. L. (2022). Robust ensemble method for short-term traffic flow prediction. *Future Generation Computer Systems-The International Journal of eScience*, 133, 395-410. https://doi.org/10.1016/j.future.2022.03.034

[10] Kumar, M., Kumar, C., Kumar, N., Kavitha, S. (2024). Efficient hotel rating prediction from reviews using ensemble learning technique. *Wireless Personal Communications*, 137(2), 1161-1187. https://doi.org/10.1007/S11277-024-11457-W

[11] Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, 35(2), 757-774. https://doi.org/10.1016/j.jksuci.2023.01.014

[12] Lin, W., Lin, Q. Z., Feng, L., & Tan, K. C. (2024). Ensemble of domain adaptation-based knowledge transfer for evolutionary multitasking. *IEEE Transactions on Evolutionary Computation*, 28(2), 388-402. https://doi.org/10.1109/TEVC.2023.3259067

[13] Taherkhani, L., Daneshvar, A., Khalili, H. A., Sanaei, M. R. (2024). Intelligent decision support system using nested ensemble approach for customer churn in the hotel industry. *Journal of Business Analytics*, 7(2), 83-93. https://doi.org/10.1080/2573234X.2023.2281317

[14] Mim, A. M., Majadi, N., & Mazumder, P. (2024). A soft voting ensemble learning approach for credit card fraud detection. *Heliyon*, 10(3), e25466. https://doi.org/10.1016/J.HELIYON.2024.E25466

[15] Guan, J. H., Yu, Z., Liao, Y. G., Tang R. B. Duan M. (2024). Predicting critical path of labor dispute resolution in legal domain by machine learning models based on shapley additive explanations and soft voting strategy. *Mathematics*, 12(2). 272 https://doi.org/10.3390/MATH12020272

[16] Gao, G., Wang H. X., & Gao, P. B. (2021). Establishing a credit risk evaluation system for smes using the soft voting fusion model. *Risks*, 9(11), 202. https://doi.org/10.3390/RISKS9110202