Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 11, No. 2, June 2025, pp. 184-193
ISSN: 2338-3070, DOI: 10.26555/jiteki.v11i2.30631

Performance Analysis of Prediction Methods on Tokyo Airbnb Data: A Comparative Study of Hyperparameter-Tuned XGBoost, ARIMA, and LSTM

Rizal Farhan Nabila Nurfalah[1], Dandy Pramana Hostiadi[2], Evi Triandini[2]
[1]Magister Program, Department of Master Information Systems, STIKOM Institute of Technology and Business, Bali, Indonesia
[2]Department of Master of Information Systems, Institute of Technology and Business STIKOM, Bali, Indonesia

ARTICLE INFO

Article history:

Received January 25, 2025
Revised March 29, 2025
Accepted April 22, 2025

Keywords:

Hyperparameter tuning;
Machine Learning;
Time Series Analysis;
Prediction Methods;
Occupancy Rate;
Optimization

ABSTRACT

The rapid growth of the digital economy has increased the importance of accurately predicting Airbnb property occupancy rates, especially in dynamic and competitive markets such as Tokyo, Japan. Property owners face significant challenges in forecasting occupancy rates due to seasonal patterns, non-linear trends, and complex temporal dependencies within the data. Addressing these challenges, this study investigates the performance of ARIMA, XGBoost, and LSTM models in predicting Airbnb occupancy rates in Tokyo. The dataset is collected from Airbnb listings and includes relevant features such as location, price, customer reviews, and historical occupancy rates. The models were optimized using Grid Search for ARIMA and Random Search for XGBoost and LSTM to identify the best hyperparameter configurations. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2), which are more appropriate for regression tasks. The results indicate that XGBoost achieves the highest R^2 (0.23), followed by LSTM (0.19) and ARIMA (0.03). However, the low R^2 values suggest that the models struggle to capture occupancy rate variations, indicating the potential influence of unmodeled external factors such as seasonality and policy changes. This study highlights the importance of hyperparameter tuning in improving prediction accuracy and contributes by providing an in-depth comparison of regression-based models for Airbnb occupancy forecasting.

Corresponding Author:

Rizal Farhan Nabila Nurfalah, Department of Institute of Technology and Business STIKOM Bali, Indonesia
E-mail: rizalfarhannn@gmail.com

1. INTRODUCTION

Digital transformation has brought about major changes in the hospitality and property rental industry, one of which is through the presence of internet-based platforms such as Airbnb [1]. The platform allows property owners to offer short-term accommodation to renters, including both international and local travelers [2]. Tokyo's dynamic rental market presents unique challenges in occupancy prediction due to seasonal variations and policy shifts. Prior research has focused on price prediction, but occupancy forecasting remains underexplored. This study bridges that gap by comparing ARIMA, XGBoost, and LSTM with hyperparameter tuning to enhance predictive accuracy. The research contributions are: (1) A comparative analysis of regression-based occupancy prediction models, and (2) An investigation of hyperparameter optimization impacts on model performance [3]. The rapid growth of the digital economy has transformed various industries, including hospitality and property rentals. Platforms like Airbnb have emerged as major players in short-term accommodation services, catering to both local and international travelers. Tokyo, Japan, as a leading global tourist destination, has a dynamic property rental market characterized by high competition influenced by seasonal patterns, local events, and shifting consumer preferences. Accurately predicting occupancy rates is crucial for property owners to optimize revenue, minimize vacancy risks, and create competitive pricing strategies [4]. However, the inherent complexity of occupancy data, including variables such as location, price, customer reviews, and historical availability, presents significant challenges in building accurate predictive models [5].

Previous studies have largely focused on different aspects of Airbnb data. For instance, Islam et al. investigated price prediction using the MESF-XGBoost approach [6], while Más-Ferrando et al. explored the factors influencing Airbnb occupancy rates during the pandemic [7]. Other research has examined regulatory impacts on short-term rentals or relied on simpler models that fail to capture non-linear trends and temporal dependencies. Despite these efforts, comparative studies that evaluate advanced prediction methods specifically ARIMA, XGBoost, and LSTM remain limited. Moreover, the role of hyperparameter optimization in enhancing model performance has not been comprehensively explored. Most previous studies have focused more on predicting Airbnb property prices or the impact of regulation on the short-term rental market. For example, a study by Islam et al. used the MESF-XGBoost approach to predict listing prices [6]. However, an in-depth comparative study of occupancy rate prediction methods, especially those involving the XGBoost approach, is [8], ARIMA [9], and LSTM [10], is still rarely done. In addition, the influence of hyperparameter tuning on model performance has not been thoroughly explored.

This study makes significant contributions by addressing these gaps. First, it provides a comparative analysis of ARIMA, XGBoost, and LSTM models, offering insights into their ability to predict Airbnb occupancy rates with diverse data characteristics. Second, it emphasizes the importance of hyperparameter tuning (using Grid Search for ARIMA and Random Search for XGBoost and LSTM) to improve predictive accuracy. Third, by focusing on Tokyo's competitive and unique property market, this study highlights context-specific applications of predictive modeling. Finally, the findings offer practical guidance for property owners and platform developers, enabling data-driven decision-making to optimize occupancy and revenue strategies. These contributions establish a foundation for future research, particularly in integrating external factors or exploring ensemble methods to further enhance predictive performance.

2. METHODS

This study uses ARIMA, XGBoost, and LSTM methods [8], [11], [12] which will later go through several important processes in making predictions for availability in a property, there are several stages that must be passed, namely, data collection, data preprocessing, data transformation, data sharing. For an overview of all stages of the process of the method used, here is a general description of the flowchart from Fig. 1. Continued by developing each model, after development, the model will be optimized using hyperparameter optimization techniques such as random search and grid search. After this stage is completed, it will be continued by evaluating each model [13].

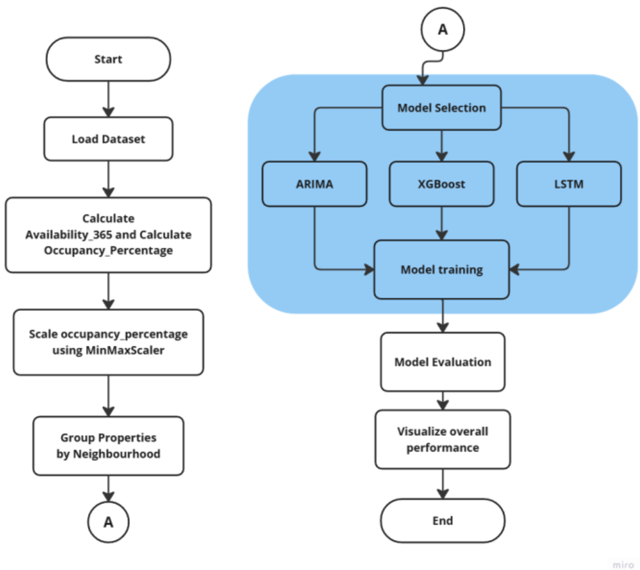


Fig. 1. Research Flowchart

2.1. Data Collection and Preparation

Table 1, the dataset used is obtained from Inside Airbnb, covering rich variables such as location, price, customer reviews, and property availability. The data with a total of 18,736 will later be separated into two data, namely with the provision of 80% for training and 20% for testing. The dataset obtained in this article covers the Tokyo area, Japan [14] from dataset public.

Table 1. Dataset Description

Variables	Description	Data Type
id	A unique identifier for each property or listing on Airbnb.	int64
name	The name of the listing or property registered on Airbnb.	object
host_id	Unique identifier for each listing owner or host.	int64
host_name	The name of the owner or host of the listing.	object
neighbourhood_group	The area/district where the property is located (all values lost).	float64
neighbourhood	The specific region where the listing is located.	object
latitude	Geographic coordinates (latitude) of the property.	float64
longitude	Geographic coordinates (longitude) of the property.	float64
room_type	Types of rooms available for rent (e.g. entire home, private room).	object
price	Property rental price per night.	float64
minimum_nights	Minimum number of nights to book.	int64
number_of_reviews	Total number of reviews the listing received.	int64
last_review	The date of the last review the listing received.	object
reviews_per_month	Average number of reviews per month.	float64
calculated_host_listings_count	The number of listings owned by the host.	int64
availability_365	The number of days in a year when the property is available for rent.	int64
number_of_reviews_ltm	Number of reviews received in the last 12 months.	int64

Table 1 this dataset is designed to provide a comprehensive picture of properties listed on Airbnb, including information about location, owner, property type, and rental activity. Each property is uniquely identified by the id variable, while name represents the name or short description that owners typically use to attract potential renters. The host or owner of the property is also identified by the host_id and host_name, which provide unique information about who manages the property. In terms of location, the dataset includes a neighbourhood variable that indicates the specific area where the property is located, as well as latitude and longitude, which capture the geographic coordinates of the property. These variables are useful for location-based analysis or data visualization using maps. While there is a neighbourhood_group column that is supposed to classify larger regions, all values in this column are empty, so they do not directly contribute to the analysis.

The type of accommodation is recorded in the room_type variable, which indicates whether the property being rented is an entire home/apt or just a private room [15], [16]. The nightly rental price is shown in price, while the minimum number of nights to book is set in minimum_nights. This information provides important insights for property market analysis or adjusting pricing strategies. Rental activity and the popularity of a property are reflected in variables such as number_of_reviews (total number of reviews), last_review (last review date), and reviews_per_month (average reviews received each month). Additionally, number_of_reviews_ltm provides insight into reviews received in the last 12 months, illustrating the current performance of a listing.

From the owner perspective, the calculated_host_listings_count variable indicates the number of other properties a host owns, giving an indication of whether the host is a casual individual or a professional host. The availability_365 variable records the number of days in a year that a property is available for rent, which is useful for assessing the availability level of a property [17]. Finally, the license column stores information about the licenses or permits held by the property, although some values in this column are blank. With a wealth of information covering location, price, rental activity, and availability, this dataset provides a rich foundation for conducting in-depth analyses of the Airbnb property market. A detailed explanation of each variable makes it easy to understand the relationship between these factors and how they impact occupancy and listing performance on the Airbnb platform.

2.2. Data Preprocessing

The data preprocessing stage aims to clean and prepare data so that it is ready to be used in analysis or machine learning models [13]. The initial steps involve data cleaning, such as handling missing values with imputation (mean, median, or mode), removing duplicates, and managing outliers using statistical techniques to reduce bias [18]. The data is then transformed, including normalization on numeric columns to have a uniform scale, as well as encoding categorical variables such as room type into numeric format [19]. Next, feature engineering is performed to create new, more meaningful features, such as location zones based on geographic coordinates. Irrelevant data is filtered out, and the dataset is divided into training (80%) and testing (20%) data to improve the model's ability to predict new data. Finally, initial visualizations such as histograms and scatter plots are used to understand data patterns, distributions, and detect anomalies, ensuring the data is in optimal condition for further analysis [20].

2.3. Classification Method

2.3.1. ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a classical statistical method used to analyze and predict time series data [9]. The implementation process begins with loading the dataset, calculating daily availability, and determining the occupancy percentage [21]. Data is grouped by region, and regions with less than 10 data are skipped. Model optimization is performed using Grid Search, which aims to find the best combination of parameters (p, d, and q) to improve prediction performance [11], [22], [23].

$$ARIMA(p,d,q) : Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-} - \cdots \theta_q \varepsilon_{t-q} \quad (1)$$

2.3.2. LSTM (Long Short-Term Memory)

LSTM is a type of artificial neural network designed to handle long-term dependencies in time series data [24]. In this study, the dataset was scaled using MinMaxScaler to ensure the data was in the appropriate scale. The data was grouped by region, and only regions with sufficient data were analyzed further [25], [26], [11]. Model hyperparameters are optimized through an efficient parameter space search. LSTMs are trained with optimal parameters, and model performance is evaluated using various metrics. LSTMs excel at learning complex patterns in time series data by preserving important information through a “gating” mechanism in their structure [27]. The general formulas that normaly uses for research is using:

$$x_t, h_{t-1}, C_{t-1} \rightarrow h_t C_t \quad (2)$$

2.3.3. XGBoost (Extreme Gradient Boosting)

XGBoost is a decision tree-based algorithm designed to handle tabular data with high efficiency. The process begins with calculating the occupancy percentage, followed by data normalization to improve learning efficiency [28], [29]. Data is grouped by region, with a certain minimum data size as an eligibility criterion. Parameter optimization is performed using Random Search to find the ideal parameter combination [8], [30]. After training with optimal parameters, the model is tested, and the evaluation results are visualized for further analysis. XGBoost offers high speed and good prediction performance on big data [6]. XGBoost itself have formula that usually use for prediction:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \quad (3)$$

2.3.4. Training Model

After data preprocessing, the dataset is split into training (80%) and testing (20%) data to ensure that the model can learn from the training data and be tested on previously unseen data, improving generalization ability [31], [32]. Three main models are used: ARIMA, LSTM, and XGBoost, each with a unique approach to handling the data:

1. ARIMAused to analyze time series data. The training process involves determining the parameters p, d, and q using the Grid Search technique. The model is trained on data that has been grouped by geographic region, with a focus on seasonal patterns and trends in occupancy rates [13], [33], [34].
2. LSTM, an artificial neural network for time series data, trained using a dataset that has been normalized with MinMaxScaler. This model utilizes a long-term memory mechanism to capture complex relationships in the data. The hyperparameter optimization process is performed with Random Search, helping to determine the best configuration, such as the number of layers, batch size, and activation function [35], [36], [37].
3. XGBoost, a decision tree-based algorithm, trained on tabular data including daily occupancy information. The model hyperparameters, including learning rate and max depth, are optimized through Random Search. This model is very efficient in handling large and complex data [38], [39], [40].

After training, the model performance is evaluated using metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R-squared [20], [41], [42], [43]. The evaluation process aims to measure the accuracy of predictions and the ability of the model to capture data patterns. Visualization of the evaluation results provides additional insight into the performance of each model, helping in selecting the best model for a given scenario.

2.3.5. Model Evaluation

The model evaluation in this study aims to assess the performance of ARIMA, LSTM, and XGBoost methods in predicting the occupancy rate of Airbnb properties [44]. The evaluation was carried out using key metrics such as Mean Absolute Error (MAE).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4}$$

Formula (4), which measures the average absolute error between the predicted and actual values. Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{5}$$

Formula (5), which places greater emphasis on large errors by calculating the square root of the mean of the errors, and R-squared (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{6}$$

Formula (6), which measures how well a model can explain variation in the data, with values close to 1 indicating a very good model. Providing a balanced view of the correct positive predictions and the model's ability to detect all positive cases [41]. The evaluation results show that ARIMA has poor initial performance with high MAE and negative R^2 values, but after optimization through Grid Search, its performance improves significantly. However, ARIMA is still less able to handle complex data compared to LSTM and XGBoost [12], [24].

3. RESULTS AND DISCUSSION

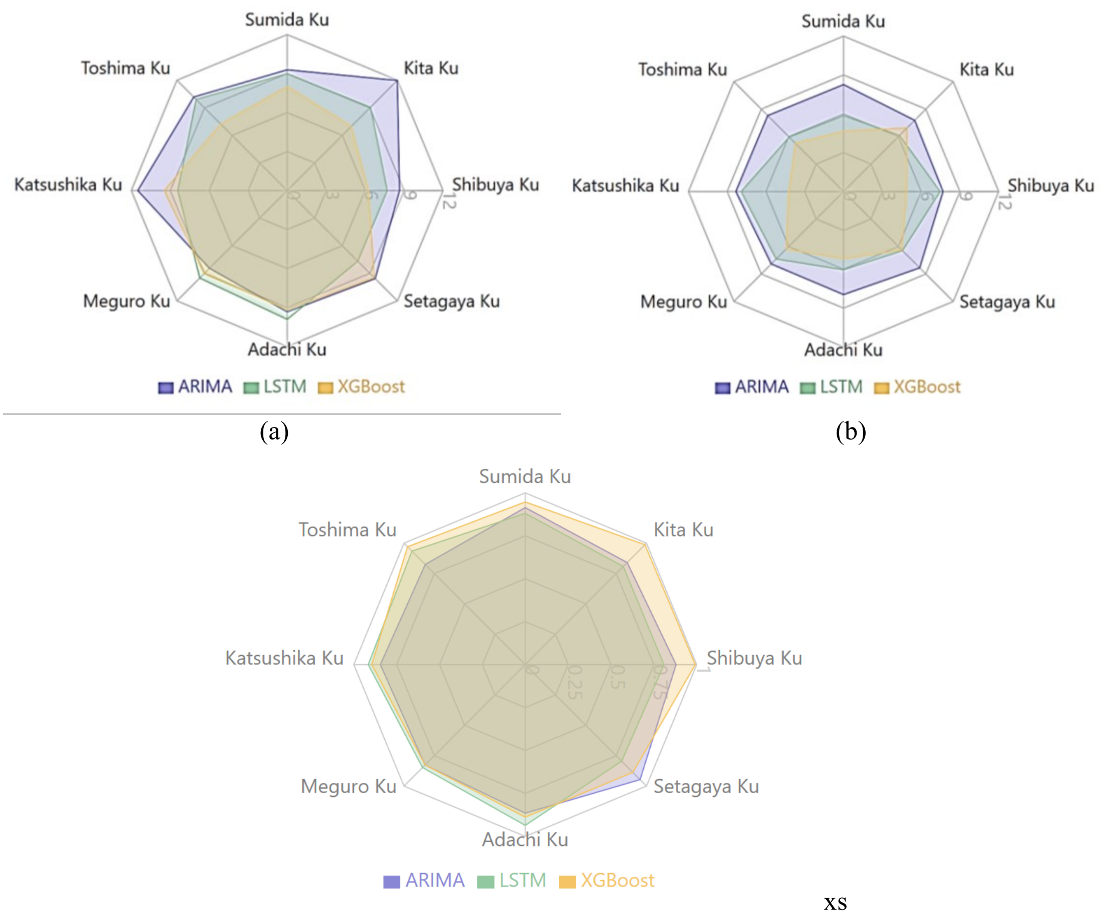
3.1. Results

The results of this study show the performance and advantages of each prediction method (ARIMA, LSTM, and XGBoost) in modeling the occupancy rate of Airbnb properties. By using evaluation metrics such as MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R^2 (R-squared), this study provides a comprehensive evaluation of the performance of each model.

Fig. 2, In the ARIMA model, performance without optimization showed less than satisfactory results, with a negative R^2 value (-0.1165), MAE of 13.0093, and RMSE of 18.174. After optimization through Grid Search, the R^2 value improved to 0.1256, with a significant reduction in MAE to 0.1124 and RMSE to 0.1587. However, despite this improvement, ARIMA remains less competitive than other methods in capturing complex data patterns.

For the LSTM model, initial results were slightly better compared to ARIMA, with an R^2 of 0.1076, MAE of 10.9591, and RMSE of 16.282. After hyperparameter optimization using Random Search, performance improved significantly, marked by a decrease in MAE to 0.0901, RMSE to 0.1256, and an increase in R^2 to 0.2534. This demonstrates that LSTM can capture non-linear patterns and complex dependencies in time series data more effectively than ARIMA.

The XGBoost model initially showed suboptimal performance, with a negative R^2 (-0.0026), MAE of 11.6442, and RMSE of 17.258. However, after optimization using Random Search, this model exhibited significant improvement, reducing MAE to 0.0823, RMSE to 0.1142, and achieving the highest R^2 value of 0.2845. This suggests that XGBoost, after tuning, is the most effective method in explaining variations in Airbnb occupancy data.



(c)

Fig. 2. Airbnb Listings and Model Performance Analysis (a) RMSE (b) MAE (c) R^2 Score

Based on Table 2, overall, the results of this study confirm the importance of parameter optimization to improve the performance of prediction models. LSTM stands out as a reliable method for handling complex and non-linear data patterns, while XGBoost excels in efficiency and prediction accuracy after optimization. On the other hand, ARIMA remains relevant for simple seasonal patterns, but has significant limitations in capturing the complexity of property occupancy data.

Table 2. Result of each method, with optimation and without optimation

Model	MAE	RMSE	R^2
ARIMA	13,0093	18,174	-0,1165
ARIMA Grid Search	0,1124	0,1587	0,1256
LSTM	10,9591	16,282	0,1076
LSTM Random Search	0,0901	0,1256	0,2534
XGBoost	11,6442	17,2580	-0,0026
XGBoost Random Search	0,0823	0,1142	0,2845

3.2. Discussion

This study aims to evaluate the performance of three Airbnb property occupancy rate prediction methods, namely XGBoost, LSTM, and ARIMA, with a focus on the effectiveness and efficiency of each method [45], [46]. The results obtained show that each model has specific advantages based on the characteristics of the data used and the purpose of the prediction. XGBoost shows the best performance in terms of efficiency and accuracy on large tabular data. With hyperparameter optimization using the Random Search technique, this model successfully achieves the highest R^2 value, demonstrating its ability to comprehensively explain data variations. In addition, the lower computation time compared to other methods makes XGBoost an ideal choice for fast applications with big data-based prediction needs [47], [48].

In contrast, LSTM excels in learning non-linear patterns and capturing long-term temporal relationships in time series data. After optimization, the model yields a significant decrease in MAE and RMSE, indicating better prediction accuracy compared to ARIMA. With its flexibility, LSTM is a great choice for complex data that requires a deep understanding of temporal patterns. Meanwhile, ARIMA provides adequate results for time series data with seasonal patterns. Parameter optimization through Grid Search shows improved performance, although ARIMA has limitations in handling complex data and non-linear patterns. This makes ARIMA more suitable for simple applications or data with stable trend patterns.

Fig. 3, LSTM, with its ability to capture non-linear patterns and long-term relationships, shows significant performance improvements after hyperparameter optimization using Random Search [43]. This model produces lower MAE and better R^2 values compared to ARIMA. Meanwhile, XGBoost, which initially showed lower performance than LSTM, managed to achieve the highest R^2 value after parameter optimization, proving its efficiency in handling large and complex data [38], [39].

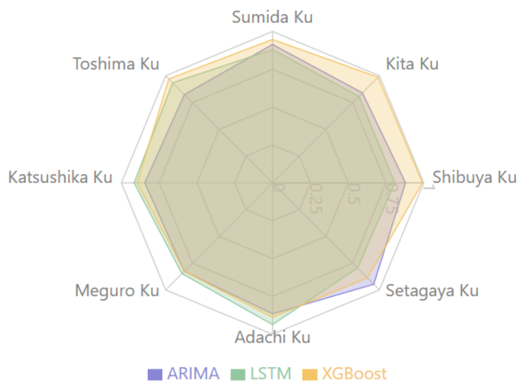


Fig. 3. Model Comparison

This study also emphasizes the importance of hyperparameter optimization in improving model performance. Optimization techniques such as Random Search for LSTM and XGBoost, and Grid Search for ARIMA, have been shown to significantly improve model performance. In addition, the use of evaluation metrics such as MAE, RMSE, and R^2 provides a clear and measurable view of the effectiveness of each method. However, there are some limitations in this study. External factors, such as seasonal trends, government policies, or major events that may affect occupancy rates, are not included in the analysis. These limitations may affect the generalization ability of the model, especially in scenarios outside the context of the dataset used. Therefore, further research is needed to integrate these external factors to improve prediction accuracy.

With the results obtained, this study makes an important contribution to the literature on occupancy rate prediction, particularly in the context of the Airbnb platform [49]. The results of the study are not only relevant for developers of prediction models but also have practical applications in supporting strategic decision-making, such as pricing and marketing planning by property owners. This study opens up opportunities for further development, both through exploration of other prediction methods, such as ensemble learning, and through integration of external variables in the analysis [50], [51].

4. CONCLUSION

This study has evaluated and compared the performance of three main prediction methods, namely XGBoost, ARIMA, and LSTM in modeling the occupancy rate of Airbnb properties. The dataset used focuses on the Tokyo, Japan area covering variables such as location, price, customer reviews, and historical availability. This study aims to provide a comprehensive view of each model's ability to handle different data characteristics. The results of the study show that XGBoost excels in handling large tabular data, producing the highest R^2 value among the three methods after hyperparameter optimization using Random Search . This method stands out in computational efficiency and the ability to explain data variations significantly. Meanwhile, LSTM shows excellent performance in capturing non-linear patterns and long-term temporal relationships in time series data. After optimization, this model significantly reduces the prediction error rate (MAE and RMSE) and becomes a very suitable choice for data with more complex patterns. On the other hand, ARIMA, although it gives good results for data with simple seasonal patterns, has limitations in handling data complexity, making it less competitive than XGBoost and LSTM. This study also emphasizes the importance of hyperparameter optimization in improving model performance. Optimization techniques such as Grid Search for ARIMA and Random Search for LSTM as well as XGBoost have significantly improved prediction accuracy. The use of evaluation metrics such as MAE, RMSE, and R^2 provides clear and measurable insights into the effectiveness of each method.

However, this study has some limitations. External factors, such as seasonal trends, local policies, or major events, were not included in the analysis. This limitation opens up opportunities for further research that can integrate external variables as well as explore additional prediction methods, such as ensemble learning, to improve model accuracy. Thus, the results of this study provide an important contribution to the occupancy prediction literature and have relevant practical applications, such as supporting strategic decision-making in pricing and marketing planning for Airbnb property owners.

Acknowledgements

The authors would like to thank the STIKOM Bali Institute of Technology and Business for the research facility support provided during the process of completing this research. The authors also express their appreciation to those who have provided relevant datasets and technical resources to support the development of machine learning models in this study. This support made it possible to conduct in-depth research on data management and machine learning-based decision making in the retail sector.

REFERENCES

[1] T. Duso, C. Michelsen, M. Schaefer, and K. D. Tran, "Airbnb and rental markets: Evidence from Berlin," *Reg Sci Urban Econ*, vol. 106, May 2024, https://doi.org/10.1016/j.regsciurbeco.2024.104007.
[2] D. A. Guttentag, S. W. Litvin, and W. W. Smith, "To Airbnb or not to Airbnb: Does Airbnb feel safer than hotels during a pandemic?," *Int J Hosp Manag*, vol. 114, Sep. 2023, https://doi.org/10.1016/j.ijhm.2023.103550.
[3] M. Makkar, S. Appau, and R. W. Belk, "Value outcomes in Airbnb as a chronotopic service," *International Journal of Research in Marketing*, 2024, https://doi.org/10.1016/j.ijresmar.2024.05.008.
[4] F. N. Priambodo and A. Sihabuddin, "An Extreme Learning Machine Model Approach on Airbnb Base Price Prediction," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 179–185, 2020, https://doi.org/10.14569/IJACSA.2020.0111123.
[5] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," in *Procedia Computer Science*, pp. 806–813, 2021, https://doi.org/10.1016/j.procs.2022.01.100.
[6] M. D. Islam, B. Li, K. S. Islam, R. Ahasan, Md. R. Mia, and M. E. Haque, "Airbnb rental price modeling based on Latent Dirichlet Allocation and MESF-XGBoost composite model," *Machine Learning with Applications*, vol. 7, p. 100208, Mar. 2022, https://doi.org/10.1016/j.mlwa.2021.100208.
[7] A. Más-Ferrando, L. Moreno-Izquierdo, J. F. Perles-Ribes, and A. Rubia, "Has COVID-19 changed the factors explaining the occupancy of Airbnb accommodation? Madrid as a case study," *Journal of Destination Marketing and Management*, vol. 31, Mar. 2024, https://doi.org/10.1016/j.jdmm.2023.100837.
[8] X. Dairu and Z. Shilong, "Machine Learning Model for Sales Forecasting by Using XGBoost," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE,* pp. 480–483, Jan. 2021, https://doi.org/10.1109/ICCECE51280.2021.9342304.
[9] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM," 2019, [Online]. Available: http://arxiv.org/abs/1911.09512.
[10] S. Gopali, S. Siami-Namini, F. Abri, and A. S. Namin, "The performance of the LSTM-based code generated by Large Language Models (LLMs) in forecasting time series data," *Natural Language Processing Journal*, vol. 9, p. 100120, Dec. 2024, https://doi.org/10.1016/j.nlp.2024.100120.
[11] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA*, pp. 1394–1401, 2018, https://doi.org/10.1109/ICMLA.2018.00227.
[12] H. Oukhouya and K. El Himdi, "Comparing Machine Learning Methods—SVR, XGBoost, LSTM, and MLP— For Forecasting the Moroccan Stock Market," *In Computer Sciences & Mathematics Forum*, vol. 7, no. 1, p. 39, 2023, https://doi.org/10.3390/iocma2023-14409.
[13] N. Zougagh, A. Charkaoui, and A. Echchatbi, "Artificial intelligence hybrid models for improving forecasting accuracy," in *Procedia Computer Science*, pp. 817–822, 2021, https://doi.org/10.1016/j.procs.2021.04.013.
[14] M. W. Murray, "Inside Airbnb: Tokyo Listings Dataset," https://insideairbnb.com/tokyo/.
[15] S. Aldahmani, "Leveraging Sentiment Analysis for Tokyo Airbnb Hosts and Leveraging Sentiment Analysis for Tokyo Airbnb Hosts and Decision Makers Decision Makers." *Rochester Institute of Technology*, 2024, [Online]. Available: https://repository.rit.edu/theses.
[16] G. Milunovich and D. Nasrabadi, "Airbnb Pricing in Sydney: Predictive Modelling and Explainable AI *," *Available at SSRN 4805859*. 2024. [Online]. Available: https://ssrn.com/abstract=4805859.
[17] D. Boto-García, R. Balado-Naves, M. Mayor, and J. F. Baños-Pino, "Consumers' Demand For Operational Licencing: Evidence From Airbnb In Paris," *Ann Tour Res*, vol. 100, May 2023, https://doi.org/10.1016/j.annals.2023.103566.

[18] H. S. Le *et al.*, "Predictive model for customer satisfaction analytics in E-commerce sector using machine learning and deep learning," *International Journal of Information Management Data Insights*, vol. 4, no. 2, Nov. 2024, https://doi.org/10.1016/j.jjimei.2024.100295.
[19] A. Laukkarinen and J. Vinha, "Long-term prediction of hourly indoor air temperature using machine learning," *Energy Build*, vol. 325, Dec. 2024, https://doi.org/10.1016/j.enbuild.2024.114972.
[20] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, "Time-series forecasting of seasonal items sales using machine learning – A comparative analysis," *International Journal of Information Management Data Insights*, vol. 2, no. 1, Apr. 2022, https://doi.org/10.1016/j.jjimei.2022.100058.
[21] Md. S. Rahman, A. H. Chowdhury, and M. Amrin, "Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh," *PLOS Global Public Health*, vol. 2, no. 5, p. e0000495, 2022, https://doi.org/10.1371/journal.pgph.0000495.
[22] S. Stevenson, "A Comparison of the Forecasting Ability of ARIMA Models," *Journal of Property Investment & Finance*, vol. 25, no. 3, pp. 223-240, 2007, https://www.emerald.com/insight/content/doi/10.1108/14635780710746902/full/html.
[23] Y. Wang, "The Prediction of the Guangzhou Housing Market Based on the ARIMA Model," *International Conference on Economics, Mathematical Finance and Risk Management (EMFRM*), vol. 38, 2023, https://doi.org/10.54691/bcpbm.v38i.3937.
[24] Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction," vol. 14, no. 8, pp. 1–7, 2022, [Online]. Available: http://arxiv.org/abs/2204.02623.
[25] D. C. Yıldırım, I. H. Toroslu, and U. Fiore, "Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators," *Financial Innovation*, vol. 7, no. 1, pp. 1–36, 2021, https://doi.org/10.1186/s40854-020-00220-2.
[26] C. Kanthila, A. Boodi, A. Marszal-Pomianowska, K. Beddiar, Y. Amirat, and M. Benbouzid, "Enhanced multi-horizon occupancy prediction in smart buildings using cascaded Bi-LSTM models with integrated features," *Energy Build*, vol. 318, Sep. 2024, https://doi.org/10.1016/j.enbuild.2024.114442.
[27] T. Falatouri, F. Darbanian, P. Brandtner, and C. Udokwu, "Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM," in *Procedia Computer Science*, pp. 993–1003, 2022, https://doi.org/10.1016/j.procs.2022.01.298.
[28] H. Wijaya, D. P. Hostiadi, and E. Triandini, "Optimization XGBoost Algorithm Using Parameter Tunning in Retail Sales Prediction," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 3, Dec. 2024, https://doi.org/10.23887/janapati.v13i3.82214.
[29] S. Yao, "Use XGBOOST to Predict the Rental Based on Airbnb Open Data," 2022, https://ceur-ws.org/Vol-3150/paper5.pdf.
[30] H. Sharma, H. Harsora, and B. Ogunleye, "An Optimal House Price Prediction Algorithm: XGBoost," *Analytics*, vol. 3, no. 1, pp. 30–45, Jan. 2024, https://doi.org/10.3390/analytics3010003.
[31] A. Malki, E. S. Atlam, and I. Gad, "Machine learning approach of detecting anomalies and forecasting time-series of IoT devices," *Alexandria Engineering Journal*, vol. 61, no. 11, pp. 8973–8986, Nov. 2022, https://doi.org/10.1016/j.aej.2022.02.038.
[32] Y. Liu, Z. Pang, M. Karlsson, and S. Gong, "Anomaly detection based on machine learning in IoT-based vertical plant wall for indoor climate control," *Build Environ*, vol. 183, Oct. 2020, https://doi.org/10.1016/j.buildenv.2020.107212.
[33] K. Lee, H. Kim, and D. H. Shin, "Forecasting Short-Term Housing Transaction Volumes using Time-Series and Internet Search Queries," *KSCE Journal of Civil Engineering*, vol. 23, no. 6, pp. 2409–2416, Jun. 2019, https://doi.org/10.1007/s12205-019-1926-9.
[34] H. Kang, K. Lee, and D. H. Shin, "Short-term Forecast Model of Apartment Jeonse Prices using Search Frequencies of News Article Keywords," *KSCE Journal of Civil Engineering*, vol. 23, no. 12, pp. 4984–4991, Dec. 2019, https://doi.org/10.1007/s12205-019-5885-y.
[35] X. Zhu, "The role of hybrid models in financial decision-making: Forecasting stock prices with advanced algorithms," *Egyptian Informatics Journal*, vol. 29, Mar. 2025, https://doi.org/10.1016/j.eij.2025.100610.
[36] J. Li, Y. Huang, Y. Lu, L. Wang, Y. Ren, and R. Chen, "Sentiment Analysis Using E-Commerce Review Keyword-Generated Image with a HybridMachine Learning-BasedModel," *Computers, Materials and Continua*, vol. 80, no. 1, pp. 1581–1599, 2024, https://doi.org/10.32604/cmc.2024.052666.
[37] Z. Gui *et al.*, "LSI-LSTM: An attention-aware LSTM for real-time driving destination prediction by considering location semantics and location importance of trajectory points," *Neurocomputing*, vol. 440, pp. 72–88, Jun. 2021, https://doi.org/10.1016/j.neucom.2021.01.067.
[38] J. Luo, Z. Zhang, Y. Fu, and F. Rao, "Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms," *Results Phys*, vol. 27, Aug. 2021, https://doi.org/10.1016/j.rinp.2021.104462.
[39] M. D. Islam, B. Li, K. S. Islam, R. Ahasan, M. R. Mia, M. E. Haque," Airbnb rental price modeling based on Latent Dirichlet Allocation and MESF-XGBoost composite model," *Machine Learning with Applications*, vol. 7, p. 100208, 2022, https://www.sciencedirect.com/science/article/pii/S2666827021001043.
[40] V. van Zoest, K. Lindberg, F. El Gohary, and C. Bartusch, "Evaluating the effects of the COVID-19 pandemic on electricity consumption patterns in the residential, public, commercial and industrial sectors in Sweden," *Energy and AI*, vol. 14, Oct. 2023, https://doi.org/10.1016/j.egyai.2023.100298.

[41] Ç. Demirel, A. A. Tokuç, and A. T. Tekin, "Click prediction boosting via Bayesian hyperparameter optimization-based ensemble learning pipelines," *Intelligent Systems with Applications*, vol. 17, Feb. 2023, https://doi.org/10.1016/j.iswa.2023.200185.
[42] X. Xu and Y. Zhang, "House price forecasting with neural networks," *Intelligent Systems with Applications*, vol. 12, p. 52, 2021, https://doi.org/10.1016/j.iswa.2021.20.
[43] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House Price Prediction using Random Forest Machine Learning Technique," in *Procedia Computer Science*, pp. 806–813, 2021, https://doi.org/10.1016/j.procs.2022.01.100.
[44] M. Parzinger, L. Hanfstaengl, F. Sigg, U. Spindler, U. Wellisch, and M. Wirnsberger, "Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection — Using machine learning methods Random Forest and LASSO," *Build Environ*, vol. 223, Sep. 2022, https://doi.org/10.1016/j.buildenv.2022.109313.
[45] H. J. Escalante, "Automated Machine Learning—A Brief Review at the End of the Early Years," in *Natural Computing Series*, pp. 11–28, 2021, https://doi.org/10.1007/978-3-030-72069-8_2.
[46] Z. N. Jawad and V. Balázs, "Machine learning-driven optimization of enterprise resource planning (ERP) systems: a comprehensive review," *Beni Suef Univ J Basic Appl Sci*, vol. 13, no. 1, 2024, https://doi.org/10.1186/s43088-023-00460-y.
[47] K. Taylor-Sakyi, "Big Data: Understanding Big Data," *arXiv preprint arXiv:1601.04602*, 2016. [Online]. Available: https://www.researchgate.net/publication/291229189.
[48] S. Wang, "An interview with Shouyang Wang: research frontier of big data-driven economic and financial forecasting," *Data Science and Management*, vol. 1, no. 1, pp. 10-12, 2021, https://doi.org/10.1016/j.dsm.2021.01.001.
[49] K. Ding, Y. Niu, W. C. Choo, "The evolution of Airbnb research A systematic literature review using structural topic modeling," *Heliyon,* vol. 9, no. 6, 2023, https://www.cell.com/heliyon/fulltext/S2405-8440(23)04298-6.
[50] O. Olubusola, N. Z. Mhlongo, D. O. Daraojimba, A. O. Ajayi-Nifise, and T. Falaiye, "Machine learning in financial forecasting: A U.S. review: Exploring the advancements, challenges, and implications of AI-driven predictions in financial markets," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 1969–1984, Feb. 2024, https://doi.org/10.30574/wjarr.2024.21.2.0444.
[51] S. A. Sayed, Y. Abdel-Hamid, and H. A. Hefny, "Artificial intelligence-based traffic flow prediction: a comprehensive review," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, pp. 1–50, 2023, https://doi.org/10.1186/s43067-023-00081-6.

BIOGRAPHY OF AUTHORS

Rizal Farhan Nabila Nurfalah, received his bachelor's degree in computer systems from Telkom University in 2019. He has experience working as a Data Analyst and Data Scientist, focusing on data-driven decision-making and machine learning. His interests include Big Data and Artificial Intelligence, and he is an active contributor to data-related communities. Email: 222012004@stikom-bali.ac.id.

Dandy Pramana Hostiadi, received his bachelor's degree in computer systems from ITB STIKOM Bali in 2010 and his master's degree in engineering from Universitas Udayana in 2014. He completed his doctoral degree in network security at Institut Teknologi Sepuluh Nopember (ITS) Surabaya in 2022. His research focuses on network security, botnet detection, and data mining. He currently serves as the Director of the Postgraduate Program at ITB STIKOM Bali and is actively involved in academic research and publication. Email: dandyph@stikom-bali.ac.id.

Evi Triandini, received her bachelor's degree in Agricultural Cultivation from Brawijaya University in 1993, her master's degree in information management from the Asian Institute of Technology in 1997, and her doctorate in Computer Science from ITS Surabaya in 2018. Her research interests include information systems, software engineering, and IoT. Email: evi@stikom-bali.ac.id.