UPGRADE: Jurnal Pendidikan Teknologi Informasi
Volume 4, No. 1, August, 2026, (1-12)
E-ISSN: 3021-7076
DOI: https://doi.org/10.30812/upgrade.v4i1.6253

Enhancing Hotel Recommendation Using Multi-Criteria Neural Collaborative Filtering

Abu Tholib [1*], Fathorazi Nur Fajri[1], Ilham Saifudin[2], Hairani [3], Juvinal Ximenes Guterres[4]
[1]Universitas Nurul Jadid, Probolinggo, Indonesia
[2]Universitas Muhammadiyah Jember, Jember, Indonesia
[3]Universitas Bumigora, Mataram, Indonesia
[4]Universidade Oriental Timur Lorosae, Dili, Timor Leste
Email Korespondensi: abu@unuja.ac.id

Article Genesis: **Accepted**: 15 May 2026 **Published**: 15 July 2026

ABSTRACT: The increasing volume of hotel information on online travel platforms made hotel selection more difficult for users because decision making had to consider multiple aspects simultaneously, including value, accessibility, service, room quality, cleanliness, and sleep quality. Conventional recommendation methods often depended on overall ratings and therefore were not sufficiently capable of representing the multidimensional nature of hotel preferences. This study proposed an improved multi-criteria neural collaborative filtering (MC-NCF) model for hotel recommendation using the Bali Hotel Review dataset. The proposed model integrated user identity, hotel identity, and six structured hotel evaluation criteria to learn user preferences in a more detailed and preference-sensitive manner. The experimental design was also strengthened through a more reliable preprocessing and evaluation pipeline, including data splitting before scaling, training-based imputation for missing values, and user ranking evaluation. The model was implemented using embedding-based neural interaction learning to capture nonlinear relationships between users, hotels, and multi-criteria features. The results showed that the proposed approach achieved stable and competitive performance across testing splits of 10%, 20%, 30%, and 40%. On the original rating scale, the model produced the best Root Mean Square Error of 0.416400 and the lowest Mean Absolute Error of 0.351719. In addition, the ranking performance remained consistently high, with Normalized Discounted Cumulative Gain values ranging from 0.976540 to 0.996243. These findings demonstrated that the proposed approach provided an effective and robust solution for hotel recommendation by leveraging structured multi-criteria preference information within a neural recommendation framework.

Keyword: hotel recommendation; multi criteria recommendation; neural collaborative filtering; preference modeling.

How to Cite:
Tholib, A., Fajri, F.N., Saifudin, I., Hairani., Guterres, J.X. (2026). Enhancing Hotel Recommendation Using Multi-Criteria Neural Collaborative Filtering. *UPGRADE: Jurnal Pendidikan Teknologi Informasi*,4(1), 1-12. https://doi.org/10.30812/upgrade.v4i1.6253

INTRODUCTION

The rapid growth of online travel platforms has significantly transformed hotel selection behavior (Majeed et al., 2026; Payandenick et al., 2026). Travelers are now required to evaluate a large volume of heterogeneous information, including hotel descriptions, prices, facilities, numerical ratings, and user-generated reviews (Alghamdi, 2025). Although such information increases transparency, it also intensifies information overload and makes the decision process more complex. In the hospitality domain, this challenge is particularly critical because hotel choice is inherently multi-attribute, involving not only price considerations but also service quality, room comfort, cleanliness, accessibility, and perceived value. Therefore, recommender systems have become increasingly important as decision-support tools for filtering alternatives and improving the relevance of hotel recommendations (Liu et al., 2025; Panarto et al., 2024).

Several studies have investigated hotel recommendation systems using various computational approaches. (Chen et al., 2021) proposed a personalized hotel recommendation method based on implicit relationships and user preferences, which was effective in capturing hidden preference patterns. however, the model mainly relied on user–item interaction information and did not explicitly incorporate structured multi-criteria hotel ratings. (Xiao et al., 2022) introduced a recommendation model that combines review text features and node features within a graph neural network framework, enabling richer representation of user and item characteristics; nevertheless, the approach depends on more complex feature extraction and graph-based modeling. (Panarto et al., 2024) developed a graph-based hotel recommendation model by integrating user–item interactions and hotel characteristics, which improved relational representation in the recommendation process; however, such models may require more complex data structures and computational resources. (Liu et al., 2025) proposed an adaptive hotel recommendation system considering attribute importance and group consensus, which strengthened preference modeling in group decision-making scenarios, but the model was less focused on integrating structured multi-criteria ratings into a neural collaborative filtering framework. Therefore, although previous studies have improved hotel recommendation performance through personalization, text-based representation, graph modeling, and adaptive preference learning, the use of structured multi-criteria hotel ratings within a practical neural collaborative filtering architecture remains insufficiently explored.

Despite these advances, existing hotel recommendation studies still leave an important methodological gap. Several studies rely heavily on unstructured review text, while others require graph-based or consensus-based architectures that may increase computational complexity. In addition, structured multi-criteria ratings such as value, accessibility, service, room, cleanliness, and sleep quality have not been fully explored within a neural collaborative filtering framework for hotel recommendation (Mastan et al., 2026). Therefore, this study proposes a Multi-Criteria Neural Collaborative Filtering model that jointly learns user identity, hotel identity, and six structured evaluation criteria to improve rating prediction and ranking quality. The objective of this study is to develop and evaluate an MC-NCF model for hotel recommendation using the Bali Hotel Review Dataset.

The remainder of this paper is organized as follows. Section 2 describes the research method, including the research design, dataset, preprocessing steps, model architecture, experimental setup, and evaluation metrics. Section 3 presents the results and discussion, including quantitative findings, visual comparisons, performance interpretation, and study limitations. Finally, Section 4 concludes the paper and outlines directions for future work.

METHOD

This research used a quantitative research approach with an experimental machine learning design, in which the proposed recommendation model was trained and evaluated under multiple train-test split scenarios. The case study of this research focused on hotel recommendation in Bali, Indonesia, using the Bali Hotel Review Dataset (Alamsyah, 2023). The dataset was obtained from Mendeley Data and contains user–hotel interactions with overall ratings and six evaluation criteria.

The preprocessing stage handled each variable group based on its characteristics. User and item identifiers were converted into numerical indices through label encoding and then represented as dense latent vectors using embedding layers. These vectors were combined to form the user–item interaction, which served as the collaborative filtering component of the model. At the same time, the multi-criteria attributes value, accessibility, service, room, cleanliness, and sleep quality were processed through missing value handling, imputation, and normalization to obtain stable numerical inputs. The interaction features and the processed multi-criteria attributes were then combined in a fusion layer within the MC-NCF framework and forwarded to a multilayer perceptron to learn nonlinear relationships and predict hotel ratings. The model performance was finally evaluated using Root Mean Square Error, Mean Absolute Error, and Normalized Discounted Cumulative Gain, as illustrated in Figure 1.

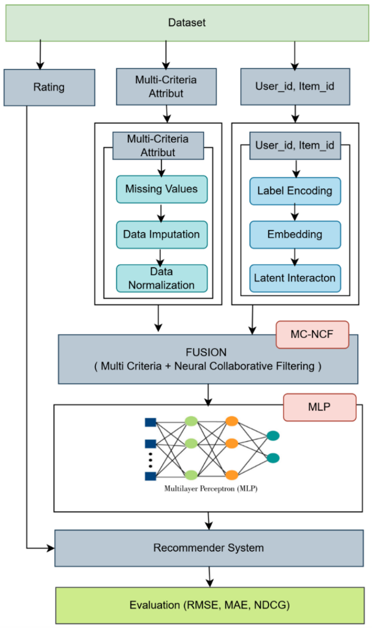

Figure 1. Proposed approach using MC-NCF

Dataset

This study utilizes the Bali Hotel Review Dataset, which contains user-generated reviews and ratings for hotels located in Bali, Indonesia (Alamsyah, 2023). The dataset provides rich information regarding user preferences toward hotels through both overall ratings and multiple evaluation criteria. Each record in the dataset represents a user–hotel interaction where a user provides ratings across several aspects of hotel services.

The presence of multiple rating criteria enables a more comprehensive representation of user preferences compared to traditional recommender systems that rely solely on a single overall rating. By incorporating these criteria, the recommendation model can capture nuanced user preferences and provide more personalized hotel recommendations. The dataset consists of thousands of interactions between users and hotels, which allows the proposed model to learn latent relationships between user behavior and hotel characteristics.

Data preparation and feature engineering

Before training the recommendation model, several preprocessing steps are performed to improve data quality and ensure efficient learning. First, data cleaning is conducted to remove missing values and ensure that all rating attributes fall within valid ranges. Any missing or inconsistent entries are filtered out to maintain data integrity. Second, user IDs and hotel IDs are encoded into numerical indices. This encoding process transforms categorical identifiers into integer representations that can be processed by embedding layers within the neural network model. Third, rating normalization is applied to the multi-criteria attributes to ensure that all rating values are within a comparable scale (Afkanpour et al., 2024; Shadbahr et al., 2023). Min–Max normalization is used to transform the ratings into a range between 0 and 1 (Li et al., 2024). This normalization helps stabilize the training process and prevents certain criteria from dominating the learning process due to scale differences.

In addition, feature engineering is performed by constructing a multi-criteria feature vector consisting of the following attributes: Value, Accessibility, Service, Room, Cleanliness, Sleep Quality. These attributes are treated as additional inputs to the recommendation model, allowing it to capture the multi-dimensional evaluation of hotels by users. Finally, the dataset was partitioned into training and testing subsets using four different ratios: 60:40%, 70:30%, 80:20%, and 90:10%. These varied split configurations were designed to provide a broader experimental setting and to evaluate the robustness, consistency, and generalization capability of the proposed model across different data partition scenarios. The training set is used to learn model parameters, while the testing set is used to evaluate the performance of the proposed recommendation model.

Models

To enhance the accuracy of hotel recommendations, this study proposed a Multi-Criteria Neural Collaborative Filtering (MC-NCF) model that integrates multi-dimensional rating criteria into the neural collaborative filtering framework. The architecture of the proposed model consists of three primary components: embedding layers, feature fusion, and a multi-layer perceptron network. First, user IDs and hotel IDs are transformed into dense vector representations using embedding layers. These embeddings capture latent characteristics of users and hotels in a low-dimensional space, allowing the model to learn hidden relationships between them. Let U represent the user embedding vector and H represent the hotel embedding vector. These vectors encode latent factors that describe user preferences and hotel attributes.

Published by Universitas Bumigora

Second, a multi-criteria rating vector is incorporated as an additional input to represent the user's evaluation of the hotel across multiple aspects. The criteria vector consists of Value, Accessibility, Service, Room, Cleanliness, and Sleep Quality. The embeddings and criteria vector are then concatenated into a unified feature representation, enabling the model to capture interactions between users, hotels, and rating criteria simultaneously. The combined feature vector is fed into a Multi-Layer Perceptron (MLP) network consisting of several fully connected layers with nonlinear activation functions (Darraz et al., 2025; Elahi et al., 2025). These layers enable the model to learn complex nonlinear relationships that cannot be captured by traditional collaborative filtering approaches (Feng and Hua, 2025; Hassanzadeh et al., 2025).

Finally, the output layer produces a predicted rating score, representing the estimated overall preference of a user toward a particular hotel (Alqahtani, 2025). To improve generalization and prevent overfitting, dropout and L2 regularization are applied during the training process (Morabbi et al., 2026; Yahia et al., 2025).

Experimental Setup

The proposed model is implemented using the Python programming language with the TensorFlow and Keras deep learning libraries. These frameworks provide efficient tools for building and training neural network-based recommender systems. The experiments are conducted using the following configuration settings Table 1.

Table 1. Hyper-Parameters Used in The Experiment For MC-NCF Models

Hyperparameter	Values
Optimizer	Adam
Learning rate	0.0001
Batch Size	256
Number of epochs	50
Embedding dimension	32
Activation function	ReLU

The Adam optimizer is selected due to its ability to adapt learning rates dynamically during training, which leads to faster convergence and improved model performance. To prevent overfitting, early stopping is applied based on validation loss, allowing the training process to terminate once the model performance stops improving. All experiments are conducted using the training dataset, while the testing dataset is used to evaluate the final performance of the proposed model.

Evaluation Metric

Evaluation metrics play a crucial role in measuring the performance of recommendation systems (Tholib et al., 2025; Priyanto et al., 2023) and determining how effective the system is at providing relevant recommendations to users (Idrissi and Zellou, 2020). This study employed several evaluation metrics, namely Root Mean Square Error (RMSE) and Normalized Discounted Cumulative Gain (NDCG). RMSE and MAE were used to measure rating prediction accuracy, where lower values indicate better predictive performance (Widiyaningtyas et al., 2025; Rezaimehr, 2024) In contrast, NDCG was used to evaluate the quality of the ranked recommendation list, where higher values indicate that the recommended items are ordered more closely to the ideal relevance ranking (Guo et al., 2023).

The formulas used in the calculation of RMSE and MAE can be found in Equation 1.

$$RMSE = \sqrt{\frac{1}{|T_{test}|} \sum_{(i,j)\in T_{test}} \left(\hat{r}_{ij} - r_{ij}\right)^2} \tag{1}$$

Where T_{test} denotes the set of user-hotel interactions in the testing data, $|T_{test}|$ represents the total number of interactions in the testing set, $\hat{r}_{ij}$ denotes the predicted rating given by user i for hotel j, and r_{ij} represents the actual rating given by user i for hotel j. The squared error term $(\hat{r}_{ij} - r_{ij})^2$ indicates the magnitude of prediction error for each user-hotel interaction.

Mean Absolute Error (MAE) is a metric used to evaluate the accuracy of a predictive model, this metric calculates the average magnitude of error in a set of predictions, without considering the direction of the error. The Equation 2 as follows:

$$MAE = \sqrt{\frac{1}{|T_{test}|} \sum_{(i,j)\in T_{test}} \left(\hat{r}_{ij} - r_{ij}\right)^2} \tag{2}$$

Where T_{test} denotes the set of user-hotel interactions in the testing data, $|T_{test}|$ is the number of testing interactions, $\hat{r}_{ij}$ represents the predicted rating of user i for hotel j, and r_{ij} represents the actual rating. The absolute error term $|\hat{r}_{ij} - r_{ij}|$ measures the absolute difference between the predicted and the actual rating.

NDCG is a metric used to assess the quality of the rankings generated by a recommendation algorithm (Mansoury et al., 2021). This metric evaluates how closely the predicted item order approximates the ideal ranking in Equation 3 and Equation 4.

$$NDCG@k = \frac{DCG@k}{IDCG@k} \quad (3)$$

$$DCG@k = \sum_{u=i}^{m} \frac{2rel_i - 1}{\log_2(i+1)} \quad (4)$$

Where K denotes the number of top recommended hotels considered in the evaluation, prepresents the position of a hotel in the recommendation list, and rel_p denotes the relevance score of the hotel at position p. $DCG@K$ measures the cumulative relevance of recommended hotels by giving higher importance to items appearing in higher positions. $IDCG@K$ represents the ideal DCG value obtained from the best possible ranking order. Finally, $NDCG@K$ normalizes the DCG value by dividing it by the ideal DCG value, producing a score between 0 and 1. A value closer to 1 indicates that the generated recommendation ranking is closer to the ideal ranking.

RESULTS AND DISCUSSION

This section presents the experimental findings of the proposed Multi-Criteria Neural Collaborative Filtering (MC-NCF) model on the Bali Hotel Review dataset. The analysis is organized into five parts, covering the experimental overview, quantitative performance, visual comparison of the training process, discussion of the findings, and reproducibility with study limitations. Through this structure, the section aims to provide a comprehensive interpretation of how the proposed model performs and how the results relate to the research problem addressed in this study.

Experimental Overview

This section presents the results of the proposed Multi-Criteria Neural Collaborative Filtering (MC-NCF) model on the Bali Hotel Review dataset. The experiment was designed to evaluate whether structured multi-criteria hotel ratings can improve both rating prediction accuracy and recommendation relevance within a neural collaborative filtering framework. In accordance with the proposed method, the model jointly learns user identity, hotel identity, and six hotel evaluation criteria, namely Value, Accessibility, Service, Room, Cleanliness, and Sleep Quality.

The final dataset used in the experiment consisted of 5,798 user–hotel interactions involving 5,390 users and 16 hotels. To evaluate robustness and generalization, the dataset was partitioned using four train-test ratios: 60:40, 70:30, 80:20 and 90:10. To ensure methodological reliability, the data preprocessing pipeline was designed carefully. Missing values in the criteria features were imputed using the mean values derived from the training set only, and Min–Max scaling was applied consistently after the data split. This design was intended to avoid data leakage and to ensure that model performance reflects genuine generalization rather than artifacts of preprocessing.

The proposed MC-NCF model was implemented using TensorFlow, which provides an efficient and flexible framework for building, training, and evaluating deep learning models with well-structured neural network components (HN et al., 2025; Nolasco Jr et al., 2025). and trained for 50 epochs with the Adam optimizer. The model used a learning rate of 0.0001, a batch size of 256, and an embedding dimension of 32. Early stopping was also employed to monitor validation loss and reduce the possibility of overfitting. The evaluation was conducted from two perspectives. First, prediction accuracy was measured using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Second, recommendation quality was assessed using user Normalized Discounted Cumulative Gain at rank 10 (NDCG@10), which measures how well the predicted ranking aligns with the ideal ranking for each user.

Quantitative Result

The quantitative results show that the proposed MC-NCF model achieved promising predictive performance on the Bali Hotel Review dataset. In the normalized rating space, the model obtained an RMSE of 0.1415 and an MAE of 0.1161. After transforming the predictions back to the original rating scale, the model achieved an RMSE of 0.4245 and an MAE of 0.3484. These values indicate that the proposed model was able to estimate users' rating preferences with relatively low prediction error.

In addition to prediction accuracy, the model also demonstrated strong ranking performance. The user evaluation produced $NDCG@10$ of 0.99624 for the evaluated users. This result indicates that, within the tested setting, the model was able to rank relevant hotel items in a manner fully consistent with the ideal ranking derived from the observed ratings. From a recommendation perspective, this is important because the practical objective of a recommender system is not only to predict ratings accurately, but also to place the most relevant items at the top of the recommendation list. Table 2 summarizes the main quantitative results obtained by the proposed model.

Table 2. Quantitative Performance of the Proposed MC-NCF Model

Metric	Ratio 60:40%	Ratio 70:30%	Ratio 80:20%	Ratio 90:10%
RMSE	0.4250	0.4198	0.4164	0.4239
MAE	0.3772	0.3760	0.3559	0.3517
NDCG	0.9962	0.9952	0.9864	0.9765

Published by Universitas Bumigora

To examine the consistency of the proposed model under different train–test partitions, additional experiments were conducted using testing proportions of 10%, 20%, 30%, and 40%. The results show that the model maintained relatively stable performance across all settings. On the original rating scale, the best RMSE was obtained at the 20% testing split with a value of 0.4164, while the lowest MAE was achieved at the 10% testing split with a value of 0.3517 and best accuracy in the NDCG on the ratio 40% testing split with a value of 0.99624, As the testing proportion increased, the prediction error showed only a modest variation, indicating that the proposed MC-NCF model remained reasonably robust under different data partition settings. Overall, these findings suggest that the model has stable predictive capability and does not exhibit substantial performance degradation when evaluated with a larger testing subset.

Figure 2. Evaluation results in terms of RMSE and MAE across different train-test ratios

Figure 1 showed that the prediction errors remained relatively stable across different train-test ratios. The RMSE values ranged from 0.41640 to 0.42501, while the MAE values ranged from 0.35171 to 0.37724. These results indicate that the proposed model maintained consistent prediction accuracy under different data partition settings.

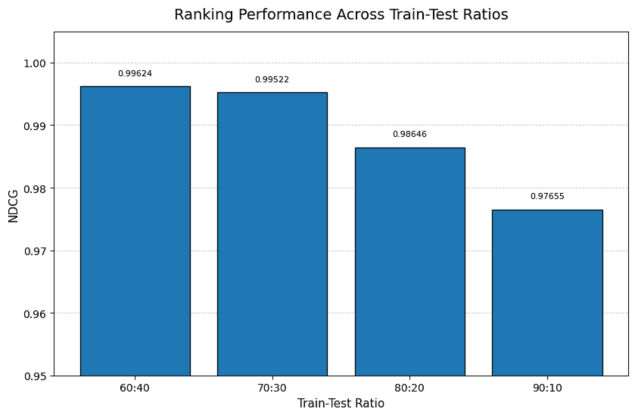

Figure 3. Evaluation results in terms of NDCG across different train-test ratios

As shown in Figure 3, showed that the ranking performance of the proposed model remained consistently high across all train-test ratios. The NDCG values ranged from 0.97655 to 0.99624, indicating that the model was effective in placing the most relevant hotel items at the top of the recommendation list. This result suggests that the proposed approach achieved strong ranking quality under different evaluation scenarios.

Published by Universitas Bumigora

Visual Comparison

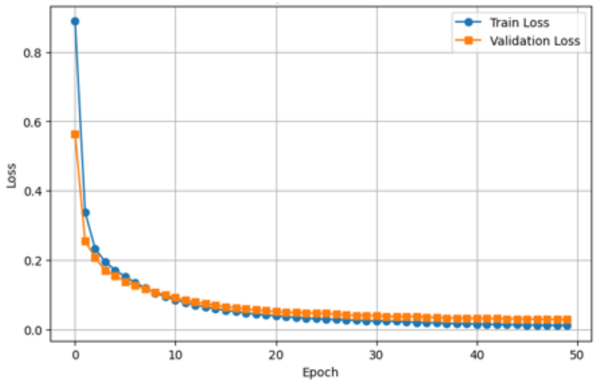

Figure 4. Training and validation loss of the proposed MC-NCF model

As illustrated in Figure 4, the training and validation loss during 50 epochs. As shown in the figure, both curves decreased sharply during the early training phase and then gradually stabilized. This pattern indicates that the model was able to learn the main structure of user preference efficiently in the initial epochs, followed by a more refined optimization process in later epochs. More importantly, the validation loss followed a decreasing trend without severe instability, which suggests that the model maintained reasonable generalization performance on unseen data.

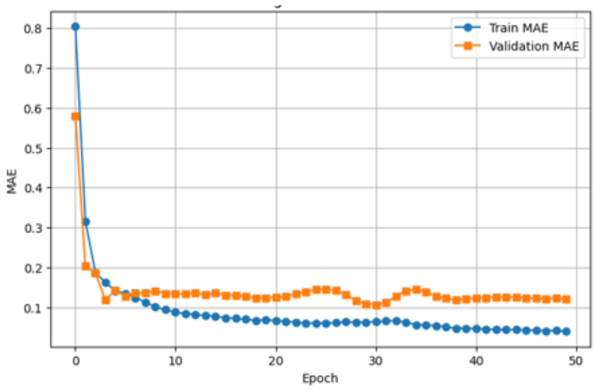

Figure 5. Training and validation MAE of the proposed MC-NCF model

As shown in Figure 5, presents the training and validation MAE curves. A pattern similar to the loss curve can be observed. The MAE values dropped substantially in the early epochs and then converged more gradually. Although the validation MAE showed minor fluctuations, it remained relatively close to the training MAE throughout the learning process. This behavior indicates that the model did not suffer from severe overfitting and that the selected hyperparameter configuration was adequate to maintain stable convergence.

Discussion

The results confirm that integrating multi-criteria hotel ratings into a neural collaborative filtering framework is a meaningful approach for hotel recommendation. In contrast to conventional recommendation models that mainly rely on overall ratings, the proposed model learns from six structured hotel evaluation criteria. This richer representation enables the model to capture user preferences in a more detailed manner and better reflects the multidimensional nature of hotel selection.

The relatively low RMSE and MAE values indicate that the proposed model was able to approximate users' overall hotel ratings with good precision. This finding is important because hotel selection is not determined by a single factor. Instead, users typically assess hotels by considering several aspects simultaneously, such as service quality, room condition, cleanliness, accessibility, and value. By incorporating these dimensions into the learning process, the proposed model is better aligned with actual user decision behavior than recommendation models based solely on single overall ratings.

The ranking performance also provides important evidence of the usefulness of the proposed approach. In recommendation systems, accurate rating prediction alone is not sufficient if the ranking of relevant items is poor. The strong $NDCG@10$ result suggests that the model was able to preserve the relative relevance order of hotel recommendations effectively. This indicates that the model is not only suitable for score estimation, but also relevant for top-N recommendation scenarios in which the placement of highly relevant hotels near the top of the list is crucial.

Another important contribution of this study lies in the experimental design. The improved pipeline avoids several common methodological weaknesses, such as applying scaling before train–test splitting, performing repeated normalization, or evaluating ranking quality without a user perspective. As a result, the reported findings are supported not only by the model architecture itself, but also by a more technically sound and reliable evaluation procedure. Therefore, the contribution of this study extends beyond model development and includes a more rigorous implementation strategy for multi-criteria recommendation experiments.

Discussion

The results confirm that integrating multi-criteria hotel ratings into a neural collaborative filtering framework is a meaningful approach for hotel recommendation. The proposed model learns from six structured hotel evaluation criteria. This richer representation enables the model to capture user preferences in a more detailed manner and better representation of the multidimensional nature of hotel selection. These findings aligned with previous studies, which demonstrate that multi-criteria ratings outperform conventional single-rating collaborative filtering and improve recommendation accuracy (Shambour et al., 2022; Zhong et al., 2021).

The relatively low RMSE and MAE values indicate that the proposed model was able to approximate users' overall hotel ratings with good precision. This finding is important because hotel selection is not determined by a single factor Instead, users typically assess hotels by considering several aspects simultaneously, such as service quality, room condition, cleanliness, accessibility, and value (Le et al., 2022). By incorporating these dimensions into the learning process, the proposed model is better aligned with actual user decision behavior than recommendation models based solely on single overall ratings. This approach addresses an important limitation of previous studies, many of which modeled user preferences using only aggregated ratings and therefore failed to exploit the richer preference information contained in attribute-specific evaluations (Le et al., 2022; Monti et al., 2021).

The ranking performance also provides important evidence of the usefulness of the proposed approach. In recommendation systems, accurate rating prediction alone is not sufficient if the ranking of relevant items is poor. The strong NDCG@10 result suggests that the model was able to preserve the relative relevance order of hotel recommendations effectively. This indicates that the model is not only suitable for score estimation, but also relevant for top-N recommendation scenarios in which the placement of highly relevant hotels near the top of the list is crucial.

Another important contribution of this study lies in the experimental design. The improved pipeline avoids several common methodological weaknesses, such as applying scaling before train–test splitting, performing repeated normalization, or evaluating ranking quality without a user perspective. As a result, the reported findings are supported not only by the model architecture itself, but also by a more technically sound and reliable evaluation procedure. Therefore, the contribution of this study extends beyond model development and includes a more rigorous implementation strategy for multi-criteria recommendation experiments.

Reproducibility and Limitation

This study was designed to support reproducibility. The experiment employed a clearly defined dataset, explicit feature construction, documented preprocessing steps, and fixed hyperparameter settings. The proposed model was implemented using widely adopted deep learning libraries, and the training process followed a transparent pipeline involving train–test splitting, missing value imputation, scaling, model fitting, and performance evaluation. In addition, the trained model and supporting encoders were stored in the notebook workflow, which facilitates reuse and replication for future experiments.

Nevertheless, several limitations should be acknowledged. First, the dataset contains only 16 hotel items, which means that the recommendation space is relatively small compared with real-world hotel booking platforms. This condition may simplify the ranking problem and may limit the generalizability of the findings to larger item spaces. Second, the ranking evaluation was conducted on a limited number of eligible users in the test data. Therefore, although the NDCG@10 result is highly encouraging, it should still be interpreted with caution. Third, the current experiment did not include a direct comparison against baseline models such as standard Neural Collaborative Filtering NCF), matrix factorization, or other multi-criteria recommendation methods. As a consequence, the current findings demonstrate the effectiveness of the proposed model internally, but do not yet establish its relative superiority in a benchmark setting.

Another limitation is that the current model relies exclusively on structured rating criteria and does not incorporate additional information such as textual reviews, temporal dynamics, travel context, or hotel metadata beyond the selected criteria. These factors may provide complementary information that could further improve recommendation quality. Future work should therefore consider comparative baseline experiments, larger and more diverse hotel datasets, and the integration of unstructured or contextual features to strengthen the robustness and generalizability of the proposed approach.

CONCLUSION

The This study addressed the limitation of hotel recommendation methods that relied primarily on overall ratings and were therefore less capable of representing multidimensional user preferences. In response to this gap, Multi-Criteria Neural Collaborative Filtering (MC-NCF) model was proposed using the Bali Hotel Review dataset. The results showed that the proposed approach achieved stable and promising performance across multiple testing splits, indicating that the integration of user identity, hotel identity, and six hotel evaluation criteria provided an effective basis for modeling hotel preferences. These findings were consistent with the research objective stated in the Introduction and confirmed that structured multi-criteria information can improve hotel recommendation performance within a neural collaborative filtering framework.

The study also showed that a more reliable preprocessing and evaluation pipeline contributed to technically sound experimental results. From an application perspective, the proposed approach has potential to support personalized hotel recommendation in digital tourism platforms. Future work may extend this study by including larger datasets, comparative baseline models, textual reviews, contextual features, and temporal interaction patterns to improve generalizability and practical deployment.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Faculty of Engineering, Informatics Engineering Study Program, University of Nurul Jadid, for the support provided in conducting this research

DECLARATION

Contributor Role Taxonomy

Author 1 (Abu Tholib): Writing-Original Draft, Writing-Review & Editing, Conceptualization, Formal Analysis, Investigation, Resources, Visualization. Author 2 and Author 3 (Fathorazi Nur Fajri and Ilham Saifudin): Conceptualization, Formal Analysis, Investigation, Supervision. Author 4 (Hairani and Juvinal Ximenes Guterres): Writing-Review & Editing, Visualization, Supervision."

Funding Statement

This research received no specific grants from public, commercial, or non-profit funding agencies.

Competing Interest Statement

The authors declare that they have no financial or personal relationships that could be perceived to influence the work reported in this article.

REFERENCES

Afkanpour, M., Hosseinzadeh, E., and Tabesh, H. (2024). Identify the most appropriate imputation method for handling missing values in clinical structured datasets: a systematic review. *BMC Medical Research Methodology*, 24(1):188. https://doi.org/10.1186/s12874-024-02310-6.

Alamsyah, A. (2023). Bali Hotel Reviews Dataset. page https://data.mendeley.com/datasets/s62ycm698z/2. https://doi.org/10.17632/s62ycm698z.2.

Alghamdi, A. (2025). Leveraging Spectral Clustering and Long Short-Term Memory Techniques for Green Hotel Recommendations in Saudi Arabia. *Sustainability*, 17(5):2328. https://doi.org/10.3390/su17052328.

Alqahtani, S. (2025). Deep-EFNet: An Optimized EfficientNetB0 Architecture With Dual Regularization for Scalable Multi-Class Brain Tumor Classification in MRI. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3567919.

Chen, K., Wang, P., and Zhang, H.-y. (2021). A novel hotel recommendation method based on personalized preferences and implicit relationships. *International Journal of Hospitality Management*, 92:102710. https://doi.org/10.1016/j.ijhm.2020.102710.

Darraz, N., Karabila, I., El-Ansari, A., Alami, N., and El Mallahi, M. (2025). Advancing recommendation systems with DeepMF and hybrid sentiment analysis: Deep learning and Lexicon-based integration. *Expert Systems with Applications*, 279:127432. https://doi.org/10.1016/j.eswa.2025.127432.

Elahi, E., Anwar, S., Al-Kfairy, M., Rodrigues, J. J. P. C., Ngueilbaye, A., Halim, Z., and Waqas, M. (2025). Graph attention-based neural collaborative filtering for item-specific recommendation system using knowledge graph. *Expert Systems with Applications*, 266:126133. https://doi.org/10.1016/j.eswa.2024.126133.

Feng, L. and Hua, W. (2025). Research on ecotourism recommendation based on graph neural network. *Procedia Computer Science*, 261:1201–1207. https://doi.org/10.1016/j.procs.2025.04.705.

Guo, L., Sun, L., Jiang, R., Luo, Y., and Zheng, X. (2023). Recommendation based on attributes and social relationships. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2023.121027.

Hassanzadeh, R., Majidnezhad, V., and Arasteh, B. (2025). A novel recommender system using light graph convolutional network and personalized knowledge-aware attention sub-network. *Scientific Reports*, 15(1):15693. https://doi.org/10.1038/s41598-025-99949-y.

HN, H. K., Muchakhandi, S., and Niranjan, P. (2025). Plant Leaf Disease Identification Using VGG, ResNet, Inception-ResNet, and DenseNet Models with Optimized Image Enhancement. pages 1–6. https://doi.org/10.1109/NMITCON65824.2025.11188183.

Idrissi, N. and Zellou, A. (2020). A systematic literature review of sparsity issues in recommender systems. *Social Network Analysis and Mining*, 10:1–23. https://doi.org/10.1007/s13278-020-0626-2.

Le, Q. H., Mau, T. N., Tansuchat, R., and Huynh, V. N. (2022). A Multi-Criteria Collaborative Filtering Approach Using Deep Learning and Dempster-Shafer Theory for Hotel Recommendations. *IEEE Access*, 10. https://doi.org/10.1109/ACCESS.2022.3165310.

Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., Ding, Y., Li, Y., Jian, L., and Cheng, J. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology*, 24(1):41. https://doi.org/10.1186/s12874-024-02173-x.

Liu, P., Dang, R., Wang, P., Xu, Y., and Zhang, Y. (2025). Online-offline combined adaptive hotel recommendation system considering attribute importance and group consensus. *Decision Support Systems*, 196:114503. https://doi.org/10.1016/j.dss.2025.114503.

Majeed, F., Siddique, A., Zafar, M., and Shaukat, K. (2026). A Fusion of Recommendation Techniques to Deliver Personalized Tourism Experience. *IET Software*, 2026(1):9218059. https://doi.org/10.1049/sfw2/9218059.

Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., and Burke, R. (2021). A Graph-Based Approach for Mitigating Multi-Sided Exposure Bias in Recommender Systems. *ACM Trans. Inf. Syst.*, 40(2). https://doi.org/10.1145/3470948.

Mastan, M., Jose, G. J. A., Rakesh, B., and Umar, S. (2026). Enhancing hotel recommendations with multi-agent reinforcement learning: a novel MADDPG approach. *Expert Systems with Applications*, page 131928. https://doi.org/10.1016/j.eswa.2026.131928.

Monti, D., Rizzo, G., and Morisio, M. (2021). A systematic literature review of multicriteria recommender systems. *Artificial Intelligence Review*, 54(1). https://doi.org/10.1007/s10462-020-09851-4.

Morabbi, S., Soltanizadeh, H., Mozaffari, S., and Fadaeieslam, M. J. (2026). Partial derivative regularized knowledge transformation for deep neural networks generalization. *Pattern Analysis and Applications*, 29(1):12. https://doi.org/10.1007/s10044-025-01559-x.

Nolasco Jr, E., Aldea, A. C., and Villaverde, J. (2025). Dog Activity Recognition Using Convolutional Neural Network. *Engineering Proceedings*, 92(1):41. https://doi.org/10.3390/engproc2025092041.

Panarto, F. A., Phanghegar, T. M., Gunardi, V., and Cenggoro, T. W. (2024). Leveraging user-item interactions and hotel characteristics for hotel recommendations in Indonesia with graph neural networks. *Procedia Computer Science*, 245:710–719. https://doi.org/10.1016/j.procs.2024.10.297.

Payandenick, M., Othman, M. K., and Payandenick, M. (2026). A multi-criteria recommendation system for personalised tourism experiences with user query analysis. *Information Technology and Tourism*, 28(1):3. https://doi.org/10.1007/s40558-025-00348-w.

Priyanto, D., Triwijoyo, B. K., Jollyta, D., Hairani, H., and Dasriani, N. G. A. (2023). Data Mining Earthquake Prediction with Multivariate Adaptive Regression Splines and Peak Ground Acceleration. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 22(3):583–592. https://doi.org/10.30812/matrik.v22i3.3061.

Rezaimehr, F. (2024). T and TRS: robust collaborative filtering recommender systems against attacks. *Multimedia Tools and Applications*, 83(11):31701–31731. https://doi.org/10.1007/s11042-023-16641-x.

Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Torne, R. V., Sala, E., and Lio, P. (2023). The impact of imputation quality on machine learning classifiers for datasets with missing values. *Communications medicine*, 3(1):139. https://doi.org/10.1038/s43856-023-00356-z.

Shambour, Q. Y., Abualhaj, M. M., Kharma, Q. M., and Taweel, F. M. (2022). A fusion multi-criteria collaborative filtering algorithm for hotel recommendations. *International Journal of Computing Science and Mathematics*, 16(4). https://doi.org/10.1504/IJCSM.2022.1005373.

Tholib, A., Widiyaningtyas, T., and Prasetya, D. D. (2025). An intelligent recommendation system utilizing a hybrid deep learning method. *Engineering, Technology & Applied Science Research*, 15(4):25971–25977. https://doi.org/10.48084/etasr.12230.

Widiyaningtyas, T., Saifudin, I., Zaeni, I. A. E., Maulana, M. Z. N., and Caesarendra, W. (2025). Memory-based collaborative filtering based on matrix factorization and Gower's set rank. *Journal of King Saud University Computer and Information Sciences*, 37(9):289. https://doi.org/10.1007/s44443-025-00261-6.

Xiao, B., Xie, X., Yang, C., and Wang, Y. (2022). RTN-GNNR: Fusing Review Text Features and Node Features for Graph Neural Network Recommendation. *IEEE Access*, 10:114165–114177. https://doi.org/10.1109/ACCESS.2022.3218882.

Yahia, A. B., Kadir, I., Abdallaoui, A., and El-Hmaidi, A. (2025). Enhancing relative humidity modelling using L2 regularization updates. *Scientific reports*, 15(1):14796. https://doi.org/10.1038/s41598-025-94356-9.

Zhong, L., Luo, Y., Zhang, X., Zhang, H., and Wang, J. (2021). Enhanced hotel recommendation method addressing the deviation between overall rating and detailed criteria ratings on Tripadvisor.com. *Journal of Intelligent and Fuzzy Systems*, 40(3). https://doi.org/10.3233/JIFS-201577.

[This page is intentionally left blank.]