International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 6, June, 2026, pp. 1787-1801
Journal homepage: http://iieta.org/journals/isi

A Hybrid Content-Based and Matrix Factorization Framework for Personalized Tourism Recommendation and Itinerary Optimization

Rifqi Hammad[1*], Apriani[2], Pahrul Irfan[3], Gede Yogi Pratama[2], Rumnan[1]

[1] Department of Software Engineering, Bumigora University, Mataram, 83127, Indonesia
[2] Department of Computer Science, Bumigora University, Mataram 83127, Indonesia
[3] Department of Informatics Engineering, Mataram University, Mataram 83125, Indonesia

Corresponding Author Email: rifqi.hammad@universitasbumigora.ac.id

https://doi.org/10.18280/isi.310602

Received: 28 July 2025
Revised: 11 September 2025
Accepted: 16 September 2025
Available online: 30 June 2026

Keywords:
personalized tourism recommendation, recommender systems, content-based filtering, matrix factorization, hybrid recommendation model, genetic algorithm, itinerary optimization

ABSTRACT

Personalized tourism recommendation remains challenging due to sparse user–item interactions, limited recommendation diversity, and the difficulty of transforming recommended destinations into practical travel itineraries. This study proposes a hybrid recommendation framework that integrates content-based filtering (CBF) and matrix factorization (MF) to improve personalized destination recommendations. It combines a genetic algorithm (GA) for itinerary optimization. Using a tourism dataset containing destination attributes and user interaction records, three recommendation strategies, including MF, CBF, and the proposed hybrid model, were systematically evaluated using precision, recall, F1-score, normalized discounted cumulative gain (nDCG), coverage, root mean square error (RMSE), and mean absolute error (MAE). The results demonstrate that MF suffers from limited recommendation effectiveness under sparse rating conditions, achieving precision, recall, F1-score, nDCG, and coverage values of 15.08%, 7.11%, 9.51%, 14.78%, and 40%, respectively. By incorporating destination content characteristics, CBF substantially improves recommendation accuracy but shows a tendency toward limited diversity. The proposed hybrid model achieves the best overall performance, with precision, recall, F1-score, nDCG, and coverage values of 59.16%, 25.98%, 35.49%, 60.95%, and 81.43%, respectively. Furthermore, the GA-based optimization module generates efficient travel sequences by considering recommended destinations and route arrangements. The proposed framework provides an integrated solution for personalized tourism recommendation and itinerary planning, demonstrating its potential for intelligent tourism platforms.

1. INTRODUCTION

Tourism is one of the most rapidly growing sectors worldwide, playing a crucial role not only in driving economic development [1-4] but also in fostering cultural preservation and promoting sustainable development [5-8]. Over the past few decades, the rise of digital technologies and artificial intelligence has transformed the way tourists plan and experience their journeys [9-12]. Recommendation systems have emerged as a key innovation in the digital era, helping travelers select destinations, optimize time, and tailor trips to personal preferences [13-15]. Nevertheless, challenges remain in designing recommendation systems that are both highly personalized and efficient, and aligned with principles of sustainability [16-18].

Indonesia, as one of the world's leading tourist destinations, reflects both the opportunities and challenges of such developments [19]. Lombok Island, located in West Nusa Tenggara (NTB), serves as a prime example, showcasing both vast tourism potential and the complexities of its management [20]. Renowned for its natural beauty and rich culture, Lombok is one of the world's top nature tourism destinations [21]. According to data from the NTB Tourism Office, Lombok hosts approximately 25 officially designated tourist attractions and 56 tourism villages [22]. This diversity highlights Lombok's appeal, while also illustrating the difficulties visitors face in selecting destinations that align with their preferences, time, and budget. The abundance of choices can often lead to decision fatigue, which may diminish the overall quality of the travel experience.

Several local applications, such as *Ayo ke Lombok* and *Lombok Trip Finder*, have been developed to provide destination information. However, their primary limitation is their insufficient capacity to generate truly personalized recommendations. Tourists are still required to conduct manual searches, with no efficient route planning support. As a result, visitors risk losing valuable time, incurring higher expenses, and experiencing suboptimal travel outcomes.

In academic literature, research on tourism recommendation systems has been extensive. Several studies have emphasized hybrid filtering approaches. Praditya et al. [23] developed a recommendation system for Palembang by combining collaborative filtering (CF) with content-based filtering (CBF). Al Fararni et al. [24] highlighted the importance of big data and artificial intelligence in building hybrid tourism recommendation models, while Garipelly et al. [25] similarly explored the integration of CBF and CF for travel systems. Forouzandeh et al. [26] applied the Artificial Bee Colony (ABC) algorithm and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to recommend top tourist destinations. Other studies by Aldayel et al. [27], Chang et al. [28], and Regula and Ali [29] also demonstrated the potential of hybrid models to enhance personalization, including leveraging social network analysis [28] and K-Nearest Neighbor algorithms [29]. Further contributions from Wang [30], Muneer et al. [31], Huda et al. [32], and Yan et al. [33] illustrate various innovative filtering approaches, ranging from preference-based to human-computer interaction-driven techniques.

Another significant strand of research emphasizes route optimization. Chen et al. [34] demonstrated the efficacy of genetic algorithm (GA) in optimizing sales routes, while Istiqfarri et al. [35] extended their use to daily vehicle routing. Lakshmi and Pallavi [36] pointed out the potential of GAs for tourism trip planning. At the same time, additional studies [37-40] have reaffirmed the relevance of genetic approaches in decision-making contexts in tourism and transportation.

From these prior studies, two major tendencies can be identified. First, hybrid recommendation systems (CBF and CF) are widely used to improve the accuracy of personalized recommendations for tourist destinations. Second, GAs have proven effective for route optimization problems. However, to date, no study has holistically integrated hybrid filtering approaches with GA-based route optimization within a single tourism recommendation framework.

Moreover, existing hybrid recommender studies primarily focus on improving recommendation accuracy by combining filtering techniques, but often overlook the optimization of post-recommendation processes, such as route sequencing and travel efficiency. Likewise, route optimization studies using GAs typically treat itinerary planning as a separate problem, disconnected from personalized user preferences. This research identifies the absence of a unified framework that integrates user-centered hybrid recommendation and genetic optimization into a coherent, intelligent tourism system.

Distinct from previous hybrid recommender models that rely solely on CF and CBF combinations, this study introduces a hybridization between CBF and matrix factorization (MF), enabling more effective feature learning from both item attributes and latent user–item interactions. The integration of GA further enhances the system by optimizing travel routes according to time, distance, and user constraints. This combined approach allows not only personalized destination recommendations but also dynamic itinerary generation that adapts to user behavior and sustainability considerations.

This study aims to develop a sustainable tourism recommendation system for Lombok Island that combines hybrid filtering for personalization with route optimization via GAs. This system is expected to deliver destination recommendations tailored to individual preferences while simultaneously suggesting more efficient itineraries. The key contribution of this research lies in the innovative integration of two previously separate approaches, producing a recommendation system that enhances the tourism experience by being personalized, resource-efficient, and consistent with sustainable tourism principles.

The key contribution of this research lies in the innovative convergence of hybrid recommendation (CBF + MF) and evolutionary optimization (GA), creating a dual-stage model that simultaneously personalizes and optimizes tourist experiences. This not only enhances the quality and contextual relevance of recommendations but also contributes to sustainable tourism practices by reducing unnecessary travel distances and resource consumption.

Accordingly, this study contributes not only to advancing knowledge in artificial intelligence-based recommendation systems but also to practical tourism applications that strengthen Lombok's competitiveness as a world-class destination. The proposed system represents a concrete step toward implementing smart tourism in Indonesia, where intelligent technologies enrich tourist experiences while safeguarding the sustainability of destinations.

2. METHODOLOGY

This study develops a hybrid recommendation system for smart tourism that integrates CBF, MF, and GAs to generate both personalized destination recommendations and optimized travel routes. The methodology is designed to address the limitations of conventional recommendation systems, which often rely solely on either content similarity or user interaction data, and typically lack route-optimization capabilities. Figure 1 illustrates the overall architecture of the proposed hybrid recommendation system for smart tourism. The system is designed with multiple integrated modules to both personalize destination recommendations and optimize travel routes. Furthermore, the proposed hybrid model is highly generalizable, enabling adaptation across various tourism domains and other recommendation datasets, thereby ensuring scalability and broad applicability.

At the top layer, the system combines CBF and CF using an MF approach to generate initial Top-k recommendations. The CBF module relies on feature extraction from destination metadata using Term Frequency–Inverse Document Frequency (TF-IDF) and on cosine similarity for similarity computation. In contrast, the CF module exploits user interaction data using MF and alternating least squares (ALSs) to predict ratings. The results from both modules are evaluated independently and then fused through a weighted hybrid approach (sequential corrections), producing the final Top-k recommendations. In addition to ranking destination relevance, the system integrates a GA module to optimize travel routes. This component applies evolutionary operators such as selection, crossover, and mutation to explore possible itineraries and identify the most efficient path connecting recommended destinations. The GA ensures that users not only receive personalized recommendations but also benefit from optimized routes tailored to their preferences and constraints.

2.1 Content-based filtering

In tourism recommendation systems, CBF leverages the textual or content features of destinations. To support this approach, data preprocessing was conducted to ensure that the collected information was clean, consistent, and suitable for each recommendation model development. The preprocessing stage involved several standardized text processing steps, including case folding, punctuation removal, and tokenization. These procedures were applied uniformly across all textual attributes to reduce noise and eliminate unnecessary variations

[41]. As a result, the preprocessed dataset was structured to facilitate subsequent feature representation. For feature extraction, the TF-IDF technique was used to represent textual attributes as vectors. TF-IDF was selected due to its effectiveness in capturing the relative importance of terms within a document while reducing the Influence of frequently occurring but less informative words, as formally defined in Eqs. (1)–(3) [42]. This transformation enabled the system to compute similarity scores between destinations and user preferences in a high-dimensional feature space.

$$tf(t,d) = \frac{\text{The number of times the term } t \text{ appears in document } d}{\text{The total number of terms present in document } d} \quad (1)$$

$$idf(t) = \log\left(\frac{1+n}{1+df(t)}\right)+1 \quad (2)$$

$$tf-idf(t,d) = tf(t,d) \times idf(t) \quad (3)$$

The term frequency $tf(t,d)$ of a word t in a document d indicates the number of times the word appears within that document, reflecting its local importance. In contrast, the inverse document frequency $idf(t)$ measures the discriminative power of a term across the entire corpus. Specifically, n denotes the total number of documents, while $df(t)$ refers to the number of documents in which the term t occurs. Meanwhile, feature engineering was performed to construct the rating matrix, which represents user–destination interactions. Each record of user visits and interactions was restructured into a matrix, with rows corresponding to users and columns to items/destinations [43]. The entries of this matrix were constructed from explicit user feedback in the form of ratings assigned to each destination they visited.

In the modeling stage, the CBF method is employed to align user preferences with destination characteristics. The textual features that have been transformed into TF-IDF vectors are compared using cosine similarity. This similarity measure computes the angular distance between two vectors in a multidimensional space, with smaller angles indicating greater similarity [44]. Mathematically, cosine similarity is computed as the ratio between the dot product of two vectors and the product of their magnitudes, as shown in Eq. (4) [45]. The resulting score lies in the range [−1, 1], with values closer to 1 indicating stronger relevance. Through this mechanism, destinations that share high semantic proximity to user preferences are prioritized in the recommendation list.

$$cosine\ similarity(A,B) = \frac{A.B}{\|A\| \ \|B\|} \quad (4)$$

where, $A.B$ denotes the dot product between A and B. $\|A\|\ \|B\|$ represents the magnitude (norm) of vectors A and B.

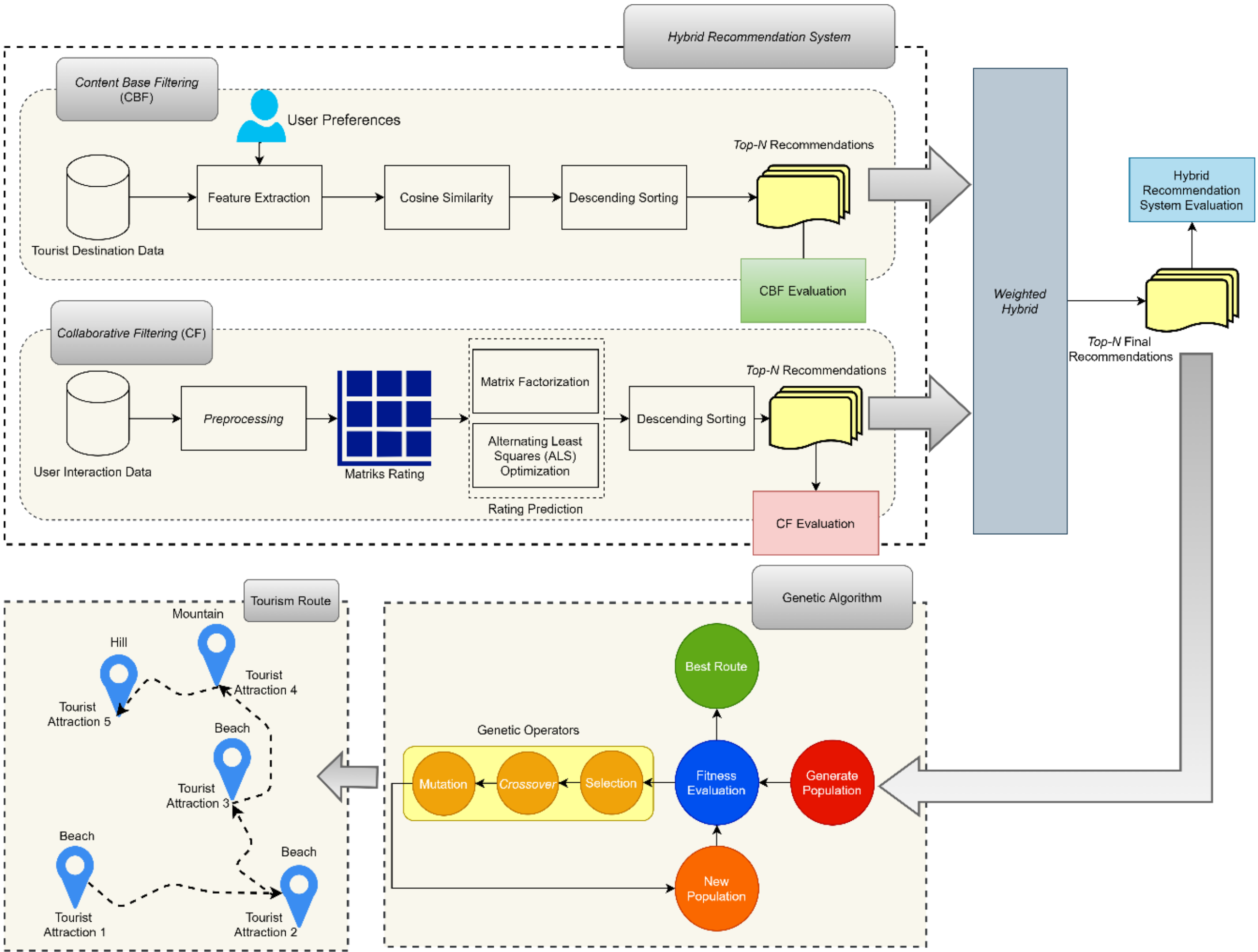

Figure 1. Proposed hybrid recommender system architecture

2.2 Matrix factorization with alternating least squares

In this study, MF was implemented using the Singular Value Decomposition (SVD) algorithm to predict missing ratings in the user–destination matrix. The user–destination matrix $R \in \mathbb{R}^{N\times M}$ was factorized into two latent matrices: the user matrix $U \in \mathbb{R}^{N\times K}$ and the item (destinations) matrix $V \in \mathbb{R}^{M\times K}$, where, k denotes the number of latent dimensions [46]. The predicted rating for a user u and item i is expressed as Eq. (5), so that the reconstructed rating matrix (user-item) can be approximated as Eq. (6):

$$\hat{r}_{ui} = U_u \cdot V_i^T \tag{5}$$

$$R \approx U \cdot V^T \tag{6}$$

MF is one of the most popular approaches in modern recommender systems due to its ability to capture latent relationships between users and items [47, 48]. The main idea is to map users and items into a lower-dimensional latent feature space so that their interactions can be represented more efficiently. In this way, the missing rating values in the user-item matrix can be predicted by multiplying two latent factor matrices, namely the user matrix U and the item matrix V. The illustration of MF is shown in Figure 2 [49].

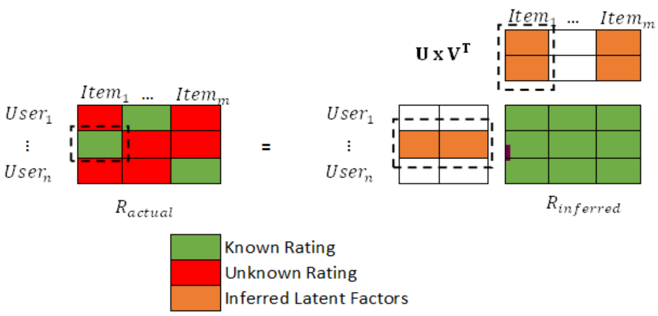

Figure 2. Illustration of matrix factorization

In the implementation, the number of latent factors was set to 20, while the training was conducted for 30 epochs with a default parameter learning rate of 0.005 and a regularization parameter of 0.02. These hyperparameters are widely used to balance model complexity and convergence speed in SVD-based MF approaches [50]. The training process can be optimized using ALSs, which iteratively minimize the reconstruction error by alternately updating user and item latent factor matrices while keeping the other fixed [51]. To evaluate the robustness of the MF model, a leave-out strategy was applied at the user level [52, 53]. First, 20% of users were randomly selected as the test set, while the remaining 80% were used as the training base. For each test user, their historical interactions (visited destinations) were split into observed and hidden subsets according to different leave-out ratios: 50%, 60%, 70%, 80%, and 90%. For example, in the 70% leave-out setting, 70% of a user's visited destinations were hidden as test records, while the remaining 30% were retained for training. The training dataset was then constructed by combining the original training users and the observed items of the test users. The MF model was trained on this dataset, and predictions were generated for the hidden items. Model performance was assessed using two categories of metrics. Rating prediction accuracy was evaluated using root mean squared error (RMSE) and mean absolute error (MAE), as formally defined in Eqs. (7)–(9) [54]. This metric quantifies the deviation between predicted and actual ratings, providing an objective measure of the model's predictive performance. Top-k recommendation quality, evaluated using Precision@5, Recall@5, and F1@5, where the top-5 recommended items for each user were compared with their hidden items.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_1 - \hat{y}_1\right)^2 \tag{7}$$

$$RMSE = \sqrt{MSE} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{9}$$

By systematically varying the leave-out ratios, this evaluation design ensures a comprehensive understanding of the MF model's ability to generalize across different levels of user information. A higher ratio (e.g., 80–90%) reflects a data-scarce condition, whereas lower ratios (e.g., 50–60%) provide more complete user histories.

2.3 Hybrid recommendation

To overcome the limitations of single-method recommendation approaches, this study employs a weighted hybrid recommendation model that integrates the outputs of both CBF and MF. Unlike traditional weighted hybrid approaches, this study adopts a sequential correction-based hybrid recommendation strategy. In this approach, the recommendation list is first fully generated by the CBF module (100% used based on ground truth), which leverages semantic similarity among destinations based on textual and numerical features computed using TF-IDF and cosine similarity. Subsequently, the recommendation list is refined using insights from MF. Specifically, destinations recommended by CBF that do not appear in the user's ground-truth interaction history are iteratively replaced with items suggested by the MF module. This mechanism ensures that the personalized, feature-driven recommendations produced by CBF are complemented with community-driven predictions from MF, thereby increasing the likelihood of matching actual user preferences. Formally, the hybrid ranking procedure can be expressed as Eqs. (10)–(13).

$$Rec_{CBF}(u) = \{i_1, i_2, \ldots, i_N\} \tag{10}$$

where, $Rec_{CBF}(u)$ denotes the initial CBF recommendations for user u, $Rec_{MF}(u)$ denotes the initial MF recommendations for user u. $Rec_{Hybrid(u)}$ denotes the final hybrid recommendations for user u.

$$Rec_{MF}(u) = \{j_1, j_2, \ldots, j_N\} \tag{11}$$

$$Rec_{Hybrid(u)} = Refine\left(Recc_{CBF}(u), Rec_{MF}(u), G(u)\right) \tag{12}$$

With the refinement operator defined as:

$$Rec_{Hybrid}(u) = \begin{cases} Rec_{CBF}(u), \ if \ Rec_{CBF}(u) \cap G(u) = Rec_{CBF}(u) \\ \left(Rec_{CBF}(u) | \ W(u)\right) \cup \left(Rec_{MF}(u) \cap G(u)\right), \ otherwise \end{cases} \tag{13}$$

Algorithm 1: Hybrid recommendation generation (content-based filtering and matrix factorization)

Input:

U**:** Set of users

I**:** Set of items (tourist destinations)

$G(u)$**:** Ground-truth visited items for user u

$Rec_{CBF}(u)$**:** CBF recommendation list for user u

$Rec_{MF}(u)$**:** MF recommendation list for user u

Output:

$Rec_{Hybrid}(u)$→: Final hybrid recommendation list for user u

Procedure hybrid recommendation (*U*, *I*):

1. For each user $u \in U$ do
2. $Rec_{CBF}(u)$ ← Generate CBF Recommendations (u, I)
3. $Rec_{MF}(u)$ ← Generate MF Recommendations (u, I)
4. $W(u)$ ← $\{ i \in Rec_{CBF}(u) | i \notin G(u) \}$ ⟶ Incorrect items in CBF
5. $Rec_{Hybrid}(u) \leftarrow Rec_{CBF}(u)$
6. For each item $i \in W(u)$ do
7. For each candidate $j \in Rec_{MF}(u)$ do
8. If ($j \in G(u)$) and ($j \notin Rec_{Hybrid}(u)$) then
9. Replace i with j in $Rec_{Hybrid}(u)$
10. Break
11. End For
12. End For
13. Return $Rec_{Hybrid}(u)$
14. End For

End Procedure

The detailed procedure of the proposed hybrid recommendation framework is presented in Algorithm 1, which illustrates how CBF outputs are sequentially refined using MF-based corrections.

Given the ground-truth visited set $G(u)$ for user u and $W(u) = \{i \in Rec_{CBF}(u) | i \notin G(u)\}$ are the "wrong" items in the CBF recommendation. Through this correction-based mechanism, the hybrid model retains the interpretability and semantic strength of CBF while incorporating MF's predictive capability to enhance accuracy. Unlike weighted schemes, where scores are combined via a parameter α, this method prioritizes CBF output as the primary recommendation and only invokes MF when corrections are required, ensuring robustness in cold-start scenarios while still leveraging collaborative knowledge.

To optimize the sequence of tourist destinations obtained from the hybrid recommendation system, this study employs a GA. GA is selected due to its effectiveness in solving combinatorial optimization problems such as the travelling salesman problem (TSP), which closely relates to the context of tourism route planning [55, 56]. Although other metaheuristic algorithms, such as Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO) are also widely applied in route optimization, GA maintains a consistent performance, proving its robustness in finding near-optimal solutions with reasonable resource utilization [57]. The GA process begins with the initialization of the population, where each individual represents a candidate route. An individual is defined as an ordered sequence of destinations $Rec_{Hybrid}(u) = \{i_1, i_2, \ldots, i_N\}$. Each individual is then evaluated using a fitness function based on the total route distance. The distance $D(Rec_{Hybrid}(u))$ is computed as the sum of geodesic distances between destinations, returning to the starting point. The fitness value is defined as shown in Eqs. (14)–(15) [58]:

$$D = \sum_{i=1}^{n-1} d\left(P_i, P_{i+1}\right) + d\left(P_n, P_1\right) \tag{14}$$

$$Fitness = \frac{1}{D} \tag{15}$$

This function is chosen for its simplicity and effectiveness in normalizing distance values into the range [0,1], ensuring that shorter routes correspond to higher fitness. In the selection phase, the algorithm uses tournament selection with tournament size $K = 3$. This method is simple yet effective in balancing exploration and exploitation compared to roulette wheel selection [59]. Several individuals are randomly chosen, and the one with the highest fitness is selected as a parent. The crossover operation adopts the order crossover (OX) technique, where a subsequence from parent 1 is directly inherited, and the remaining positions are filled from parent 2 without duplication. This method is appropriate for TSP-like problems because it preserves the relative order of destinations and ensures route validity. The mutation process is performed using swap mutation, in which two positions in the route are exchanged with a probability of mutation $p_m = 0.1$. A relatively low mutation rate (5–15%) is recommended to avoid premature convergence while maintaining population diversity [60]. At each generation, the best individual is recorded to guarantee elitism. The evolutionary process continues until the maximum number of generations is reached or the population converges. The hyperparameters used include a population size of 10 individuals and 200 generations. A relatively small population size is selected to maintain computational efficiency, while the number of generations (stopping condition) is sufficient to ensure convergence toward the optimal solutions. With this design, GA is capable of generating optimized tourism routes with minimal travel distance based on recommended destinations. Integrating GA into the hybrid recommendation system ensures that the final output is not only relevant to user preferences but also efficient for travel planning.

2.4 Evaluation

To rigorously assess the effectiveness of the proposed recommendation system, this study adopts a set of standard information retrieval metrics, namely Precision@k, Recall@k, and F1-score (k = 5). These metrics are widely used in recommender system evaluation because they capture complementary aspects of recommendation quality. Precision@k evaluates the proportion of recommended items that are relevant to the user. Recall@k measures the proportion of relevant items that are successfully retrieved within the Top-k list. While F1-score provides a harmonic mean between precision and recall to balance both perspectives, as defined in Eqs. (16)–(18). In addition, the evaluation includes normalized discounted cumulative gain (nDCG) and coverage metrics, as defined in Eqs. (19)–(21). nDCG quantifies the ranking quality by emphasizing the position of relevant items within the recommendation list, while coverage measures the diversity and distribution of recommended destinations across all users [61].

$$Precision@k(u) = \frac{|rel(u) \cap rec_k(u)|}{k} \quad (16)$$

$$Recall@k(u) = \frac{|rel(u) \cap rec_k(u)|}{rel(u)} \quad (17)$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (18)$$

$$Coverage = \frac{|\text{ Unique Items Recommended }|}{|\text{ Total Items in Catalog }|} \times 100\% \quad (19)$$

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (20)$$

$$nDCG_p = \frac{DCG_p}{iDCG_p} \quad (21)$$

In the evaluation process, each user u is identified by a unique identifier. For every user, $rel(u)$ denotes the ground truth set of relevant items obtained from the test dataset, while $rec_k(u)$ refers to the list of Top-k items recommended by the system. The parameter k specifies the cutoff value that determines how many items are included in the recommendation list. The Discounted Cumulative Gain (DCG) is calculated from the relevance scores rel_i of items at each rank i among p ranked items. The nDCG is obtained by normalizing DCG with the ideal DCG (iDCG), representing a perfect ranking, yielding a score between 0 and 1 where higher values indicate better ranking quality. Once precision, recall, F1-score, and nDCG are calculated for each user, the overall system performance is reported as the average of these metrics across all users in the test set.

3. RESULTS AND DISCUSSION

This section presents the study's results, focusing on both data preparation and the performance evaluation of the proposed recommendation models. The results are structured into two main parts: first, the dataset and preprocessing procedures are described to provide context on how the raw data were transformed into formats suitable for model training; second, the experimental results are analyzed, comparing the performance of CBF, MF, and a hybrid approach, including insights on route optimization using a GA. This structured approach allows a comprehensive understanding of both the data handling and the effectiveness of each recommendation approach.

3.1 Dataset and data preparation

The dataset employed in this study consists of 70 tourist destination records of tourist destinations enriched with attributes such as category, facilities, activities, budget, and geographical coordinates (latitude and longitude). As an initial step, the destinations' data were visualized on a map of Lombok by plotting the latitude and longitude values. Each destination was marked as a node on the map, allowing a geographical overview of the distribution of attractions across the island, as shown in Figure 3. In total, the dataset involves 958 registered users and 12,508 recorded interactions between users and destinations, which consist solely of rating activities provided by users to express their level of satisfaction or interest in each destination. Table 1 illustrates a sample of the raw destination dataset, highlighting the presence of both textual and numerical attributes. Table 2 presents a sample of user preference data, including the corresponding codes and indicated interests. This dataset serves as the basis for evaluating the recommendation system, allowing the models to match user preferences with suitable tourist destinations. The raw textual fields, such as category, facilities, and activities, contain unstructured information that requires systematic preprocessing before feature extraction. The dataset used in this study was obtained from the Lomboktripfinder.com database through structured query language (SQL)-based aggregation. Data collection followed ethical and legal requirements, with user consent and developer approval secured during registration. The data were used responsibly to support system development and enhance recommendation features.

Following preprocessing, which involved tokenization, case-folding, and removal of irrelevant tokens, the unstructured textual attributes were converted into token lists. For instance, the destination category "*Air Terjun*" was transformed into the tokenized sequence ["air", "terjun"]. Tables 3 and 4 present a sample of the preprocessed data. This transformation was critical in enabling the application of feature extraction techniques, particularly TF-IDF, which converts the tokenized text into numerical vectors. By representing destination attributes in vector space, the system captures both term importance and inter-destination similarity. Such representation provides the foundation for similarity-based methods, particularly in CBF, where cosine similarity is used to quantify the closeness between user preferences and available destinations.

Table 1. Sample data of tourist attraction

Destination Code	Destination Name	Category	Facilities	Activities	Budget	Latitude	Longitude
DST-001	Air Terjun Sendang Gile	Waterfall	Parking, Health or First Aid Services, …, Security or Guards	Hiking, Photography, Camping, Wildlife Watching	100000	-8,305 901347	116,40 83415
DST-002	Air Terjun Tiu Kelep	Waterfall	Parking, Public Restrooms, …, Tourist Information Center	Hiking, Swimming, Camping, Water Sports	100000	-8,312 57795	116,40 40733
DST-003	Air Terjun Benang Kelambu	Waterfall	Parking, Public Restrooms, …, Security or Guards	Hiking, Photography, Culinary, Swimming, Camping, Water Sports	50000	-8,532 4247	116,34 11681

Table 2. Sample data of user preferences

Preference Code	User Code	Name	Category	Facilities	Activities	Budget
PR-00009	USR-00009	Bayu Kusuma	Waterfalls, Cultural Heritage, Nature Tourism	Parking, Public Restrooms, Places of Worship, …, Tourist Information Center	Hiking, Photography, Swimming, Water Sports, ..., Cultural Tourism	210000
PR-00010	USR-00010	Rendy Firmansyah	Waterfalls, Nature Tourism	Parking, Public Restrooms, Restaurants or Cafes, …, Places to Stay	Hiking, Photography, Swimming, Camping, Water Sports, ..., Hiking	1150000
PR-00011	USR-00011	Teguh Permata	Tourist Villages, Nature Tourism	Parking, Public Restrooms, Restaurants or Cafes, …, Children's Play Area	Photography, Culinary, Camping, Cycling or ATV, ..., Wildlife Observation	550000

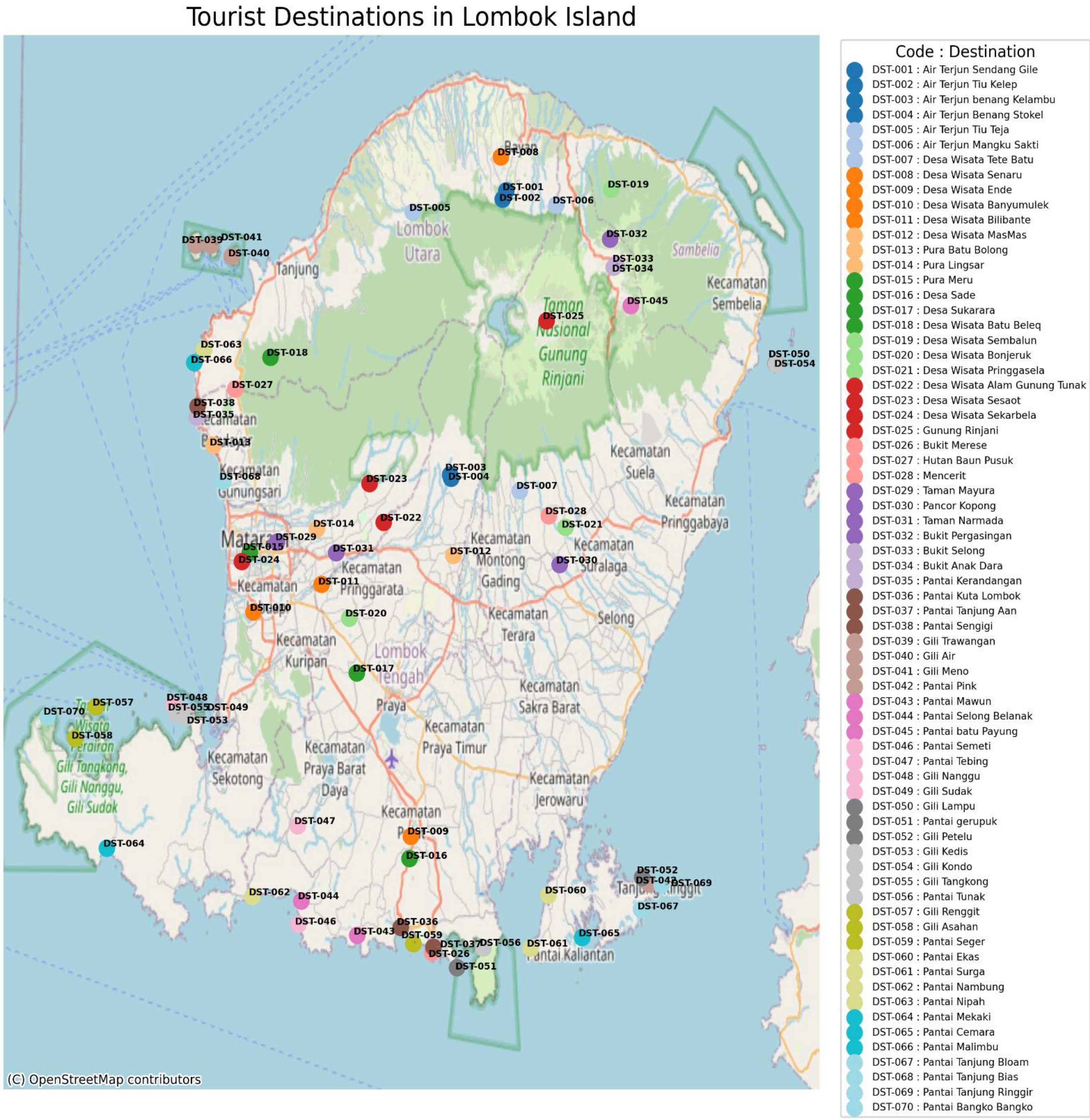

Figure 3. Geographical plot of tourist destinations in Lombok Island

Table 3. Preprocessed sample data of tourist attractions

Destination Code	Destination Name	Category	Facilities	Activities	Budget	Latitude	Longitude
DST-001	[“air”, “terjun”, “sendang”, “gile”]	["waterfall"]	[“place”, “parking”, “service”, “health”, “or”, “help”, …, “security”, “or”, “guard”]	["hiking", "photography", "camping", "observing", "animal", "wildlife"]	100000	-8,305 901347	116,40 83415
DST-002	[“air”, “terjun”, “tiu”, “kelep”]	["waterfall"]	[“place”, “parking”, “toilet”, “public”, …, “center”, “information”, “tourism”]	["hiking", "swimming", "camping", "sports", "water"]	100000	-8,312 57795	116,40 40733
DST-003	[“air”, “terjun”, “benang”, “kelambu”]	["waterfall"]	[“place”, “parking”,	["hiking", "photography", "culinary", "swimming", "camping", "sports", "water"]	50000	-8,532 4247	116,34 11681

Table 4. Preprocessed sample data of user preferences

User Code	Category	Facilities	Activities	Budget
USR-00009	["waterfall", "cultural", "heritage", "nature", "tourism"]	["parking", "public", "toilet", "place", "of", "worship", ..., "tourist", "information", "center"]	["hiking", "photography", "swimming", "sports", "water", ..., "tourism", "culture"]	210000
USR-00010	["water", "fall", "tourism", "nature"]	["parking area", "public”, “toilet", "restaurant", "or", "cafe", …, "accommodation"]	["hiking", "photography", "swimming", "camping", ..., "sports", "water"]	1150000
USR-00011	["village", "tourism", "tourism", "nature"]	[“place”, “parking”, “toilet”, “public”, “restaurant”, “or”, “cafe”, …, “area”, “play”, “children”]	["photography", "culinary", "camping", "cycling", "or" "ATV", ..., "observing", "animal", "wildlife"]	550000

Table 5. Sample user visit or interaction data

Preference Code	User Code	Destination Code	Rating
PR-00001	USR-00001	DST-004	4
PR-00001	USR-00001	DST-027	3
PR-00002	USR-00002	DST-048	4
PR-00002	USR-00002	DST-014	3
PR-00003	USR-00003	DST-010	3
PR-00003	USR-00003	DST-021	4

Table 6. Preprocessed sample user interaction data as a rating matrix

User Code / Destination Code	DST-004	DST-010	DST-014	DST-021	DST-027	DST-048
USR-00001	4	0	0	0	3	0
USR-00002	0	0	3	0	0	4
USR-00003	0	3	0	3	0	0

Table 5 presents a sample of user visits or interaction data, including the reference code, user code, destination code, and the corresponding rating provided by the user. This dataset captures explicit feedback from users regarding their experiences with different tourist destinations. To facilitate matrix-based recommendation methods, the raw interaction data were preprocessed into a rating matrix, as shown in Table 6. In this matrix, rows represent users, columns represent destinations, and each cell contains the user-assigned rating for a specific destination. A value of 0 indicates that the user has not rated or visited that destination. This transformation is essential for applying MF, as it provides a structured format that allows the models to learn latent factors representing both user preferences and item characteristics. Table 7 presents sample results of the TF-IDF feature extraction applied to the preprocessed destination attributes. The table illustrates how textual information from categories, facilities, and activities is transformed into numerical feature vectors, capturing the relative importance of each term. These vectors form the basis for computing similarity scores between user preferences and destinations in the CBF model.

Table 7. Sample results of Term Frequency–Inverse Document Frequency (TF-IDF) feature extraction

Feature Extraction	Category	Facilities	Activities	Budget
Input text	[“water”, “fall”]	["place", "parking", "service", "health", "or", "help/assistance", …, "security", "or", "guard"]	["hiking", "photography", "camping", "wildlife", "watching"]	100000
TF-IDF	array([[0.12, 0.57, 0.0, 0.0, 0.08, …, 0.03, 0.68, 0.0, 0.04]]) Vector Shape (70, 10)	array([[0.0000, 0.3127, 0.0000, 0.0000, 0.2839, 0.0000, 0.01637, 0.0000, …, 0.0000, 0.0000, 0.0000, 0.2948, 0.0000, 0.0000]]) Vector Shape (70, 37)	array([[0.1147, 0.0000, 0.0000, 0.02392, 0.0000, …, 0.3805, 0.0000]]) Vector Shape (70, 23)	100000

3.2 Experimental results and discussion

The evaluation of the MF model was conducted across multiple leave-out scenarios, ranging from 50% to 90%. Following the MF analysis, a comparative evaluation was performed between CBF, the best-performing MF scenario (precision consideration), and the hybrid model that integrates both approaches. This comparison aims to determine which method delivers the most accurate and practical recommendations for tourism. Table 8 presents the evaluation results of the MF model in the tourism recommendation system using key metrics: precision, recall, F1-score, nDCG, coverage, RMSE, and MAE. Precision emerges as the most prominent metric, showing a relatively consistent increase as the leave-out proportion grows from 50% to 90%. It is important to note that a higher leave-out percentage means more items are hidden during evaluation, giving the model more opportunities to "guess" relevant items correctly. Consequently, the observed improvement in precision with increasing leave-out is expected and reflects the model's ability to predict hidden interactions as the number of test items increases. Meanwhile, recall and F1-score remain relatively low and do not show significant changes across scenarios. Interestingly, the highest recall is observed in the leave-out 60% scenario, at 8.48%, indicating the model's enhanced ability to capture positive user interactions at this specific training data proportion.

Regarding RMSE, the prediction error does not vary significantly across all evaluation scenarios, ranging from 0.5315 to 0.5377, while the corresponding MAE values range from 0.5073 to 0.5124. This consistency indicates that both error metrics demonstrate stable performance under different proportions of training and testing data. The relatively small difference between RMSE and MAE suggests that large prediction deviations are rare, confirming the robustness of the MF model in predicting user ratings. Furthermore, the ranking performance, as measured by nDCG and coverage, also shows steady results across the evaluation settings, with average nDCG values ranging from 8.23% to 14.78% and coverage levels between 35.71% and 40%. These findings indicate that although the MF model maintains a consistent ability to rank relevant destinations, its recommendation diversity remains moderate.

In other words, changes in the amount of training data primarily affect the model's ability to recognize specific user preferences (precision and recall) but do not substantially impact the overall numerical prediction error. Figure 4 presents a bar chart comparing the precision, recall, and F1-score metrics across different leave-out scenarios. The chart shows that precision increases consistently as the leave-out proportion grows, while recall and F1-score remain low and relatively stable. This indicates that although the model performs well in accurately predicting specific positive interactions (precision), its ability to capture all relevant user interactions (recall) is still limited. Practically, these findings highlight the importance of selecting an optimal proportion of training data. The 60% leave-out scenario achieves the highest recall, suggesting a balance point between the model's ability to recognize positive user interactions and its stability in predicting ratings overall. This strategy aligns with cross-validation and regularization practices commonly applied in modern recommender systems to ensure effectiveness without overfitting to the training data.

Table 8. Performance of the matrix factorization (MF) model under different leave-out scenarios

No.	Evaluation Scenario	Average Precision	Average Recall	Average F1-Score	Average nDCG	Coverage	RMSE	MAE
1	Leave-out 50%	7.96%	6.52%	7.03%	8.23%	38.57%	0.5359	0.5092
2	Leave-out 60%	12.15%	8.48%	9.79%	12.71%	38.57%	0.5343	0.5073
3	Leave-out 70%	12.36%	7.33%	9.04%	12.58%	35.71%	0.5377	0.5119
4	Leave-out 80%	14.14%	7.41%	9.57%	13.78%	38.57%	0.5315	0.5084
5	Leave-out 90%	**15.08%**	7.11%	9.51%	14.78%	40%	0.5332	0.5124

Note: nDCG: normalized discounted cumulative gain; RMSE: root mean square error; MAE: mean absolute error.

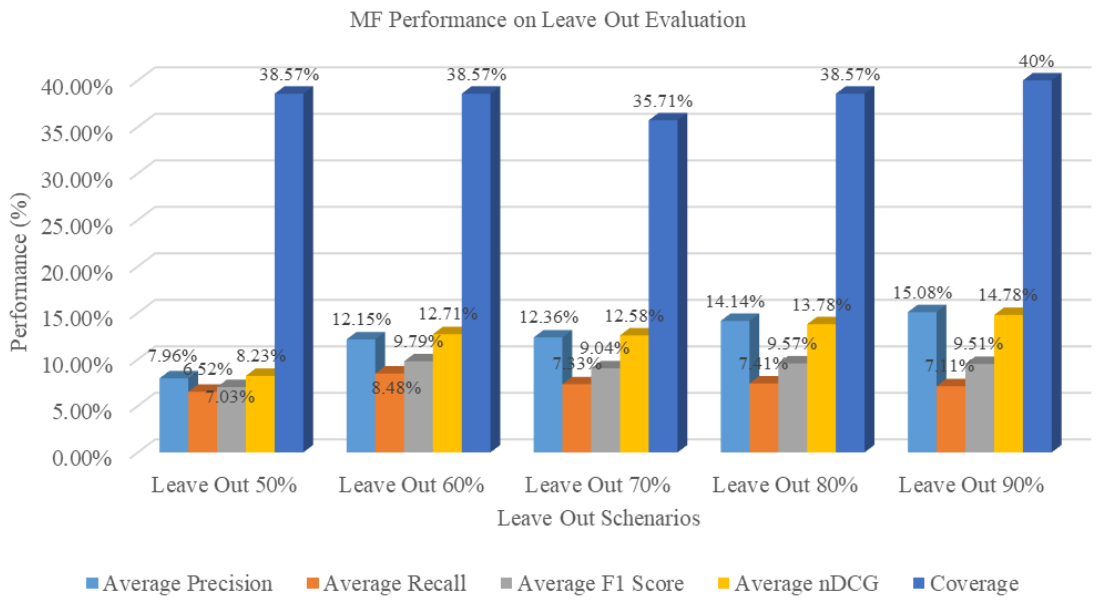

Figure 4. Bar chart comparison of precision, recall, and F1-score across leave-out scenarios (50%–90%)

Table 9. Performance comparison of content-based filtering (CBF), matrix factorization (MF) (leave-out 90%), and hybrid approaches

No.	Model Approach	Average Precision	Average Recall	Average F1-Score	Average nDCG	Coverage
1	CBF	48.69%	21.64%	29.47%	52.58%	80%
2	Best MF (leave-out 90%)	15.08%	7.11%	9.51%	14.78%	40%
3	Hybrid (CBF + MF)	**59.16%**	**25.98%**	**35.49%**	**60.95%**	**81.43%**

Note: nDCG: normalized discounted cumulative gain.

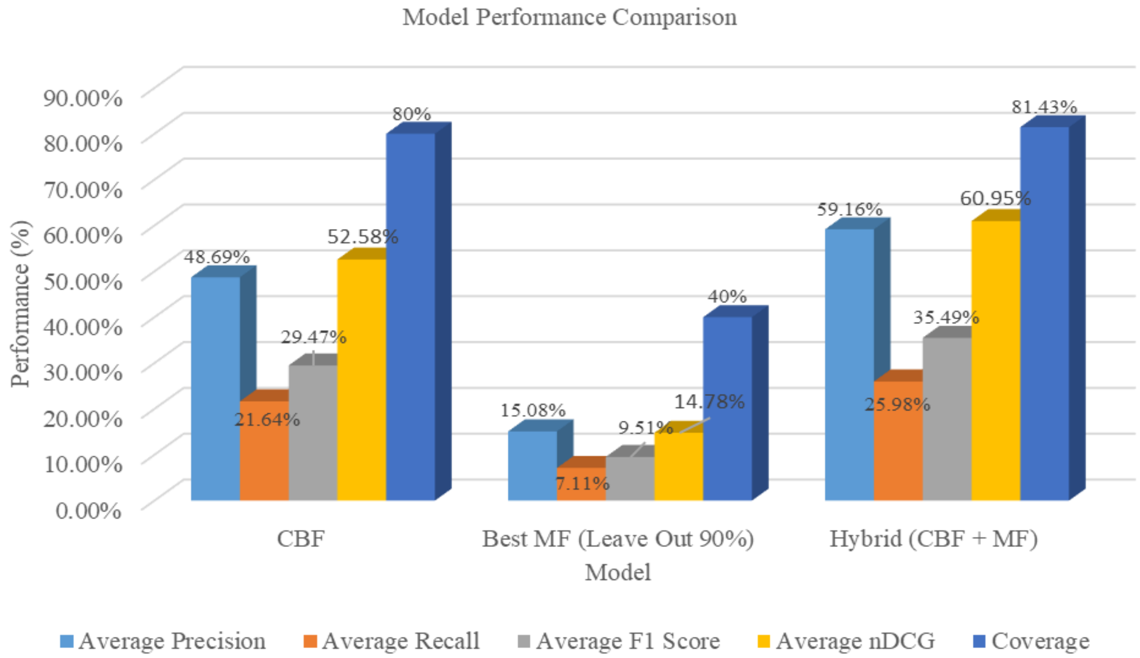

Figure 5. Performance comparison of content-based filtering (CBF), matrix factorization (MF), and hybrid models

Table 9 presents a performance comparison of three approaches: CBF, MF under the 90% leave-out scenario as the best MF model based on precision, and the hybrid (CBF + MF). Table 9 also shows that CBF outperforms MF, achieving a precision of 48.69%, a recall of 21.64%, an F1-score of 29.47%, nDCG of 52.58%, and coverage of 80%, while MF yields lower values with a precision of 15.08%, a recall of 7.11%, an F1-score of 9.51%, nDCG of 14.78%, and coverage of 40%. The weaker performance of MF is likely due to the limited amount of user interaction data available in the 90% leave-out scenario, making it difficult for the model to capture preference patterns [62] fully. MF heavily relies on sufficient user–item interaction data to learn latent factors, whereas CBF leverages complete destination content information, enabling it to still provide more relevant recommendations despite sparse interaction data. Combining both approaches in the hybrid (CBF + MF) leads to a significant improvement across all metrics: precision rises to 59.16%, recall to 25.98%, F1-score to 35.49%, nDCG of 60.95%, and coverage of 81.43%. This confirms that integrating destination content information with user interaction patterns enhances recommendation quality, as the hybrid model exploits the strengths of both methods. These results are consistent with recent studies showing that hybrid models consistently outperform a standalone approach in recommendation systems across various domains [63]. Practically, the findings highlight the importance of hybrid approaches for tourism recommendation systems, where combining destination content and user preferences yields more accurate and comprehensive recommendations, ultimately improving user experience and fostering greater engagement with the tourism platform.

To provide further clarity, Figure 5 presents a bar chart comparing the performance of three recommendation approaches: CBF, MF (leave-out 90%), and hybrid. The results show that the hybrid consistently achieves the highest performance across all metrics, followed by CBF, while MF performs the lowest. Although MF maintains a stable RMSE in rating prediction, its precision and recall remain relatively low, reflecting the limitations of relying solely on sparse user interaction data. CBF, on the other hand, effectively captures user preferences through content information and achieves higher precision, but it is prone to overspecialization, where recommendations become too similar to previously liked items, limiting diversity and exploration of new preferences [64]. This is reflected in its relatively low recall, as CBF struggles to capture more diverse or hidden user interests. The hybrid approach successfully addresses these limitations by combining the strengths of CBF and MF, resulting in both higher precision and recall. By leveraging destination content and user interaction patterns, hybrid produces more relevant, diverse, and comprehensive recommendations while reducing the overspecialization risks inherent in pure content-based methods. In terms of ranking quality, the nDCG results indicate that CBF performs substantially better than MF, while the hybrid model achieves the highest score, reflecting superior ranking relevance. Similarly, Coverage improves from 40% in MF to 80% in CBF and slightly higher at 81.43% in the hybrid model, demonstrating its capability to balance accuracy and diversity.

Table 10. Sample recommendation results generated by content-based filtering (CBF)

User Code	Ground Truth	Recommendation	Hits	Precision@5	Recall@5	F1@5	nDCG@5
USR-00031	['DST-032', 'DST-030', 'DST-025', 'DST-028', 'DST-036', 'DST-001', 'DST-065', 'DST-054', 'DST-037', 'DST-033', 'DST-042']	['DST-006', 'DST-048', 'DST-005', 'DST-054', 'DST-040']	['DST-054']	20%	9.09%	12.5%	14.60%
USR-00032	['DST-014', 'DST-013', 'DST-056', 'DST-070', 'DST-036', 'DST-060', 'DST-053', 'DST-063', 'DST-066', 'DST-040', 'DST-004', 'DST-062', 'DST-041', 'DST-025', 'DST-055', 'DST-064', 'DST-069', 'DST-054', 'DST-037']	['DST-050', 'DST-041', 'DST-048', 'DST-053', 'DST-054']	['DST-041', 'DST-053', 'DST-054']	60%	15.78%	25%	51.47%
USR-00674	['DST-044', 'DST-041', 'DST-048', 'DST-064', 'DST-036', 'DST-035', 'DST-049', 'DST-045', 'DST-070']	['DST-050', 'DST-039', 'DST-041', 'DST-048', 'DST-040']	['DST-041', 'DST-048']	40%	22.22%	28.57%	48.52%

Table 11. Sample recommendation results generated by matrix factorization (MF) 90

User Code	Ground Truth	Recommendation	Hits	Precision@5	Recall@5	F1@5	nDCG@5
USR-00031	['DST-032', 'DST-030', 'DST-025', 'DST-028', 'DST-036', 'DST-001', 'DST-065', 'DST-037', 'DST-033', 'DST-042']	['DST-050', 'DST-019', 'DST-025', 'DST-008', 'DST-003']	['DST-025']	20%	10%	13.33%	21.39%
USR-00032	['DST-014', 'DST-062', 'DST-004', 'DST-041', 'DST-064', 'DST-025', 'DST-055', 'DST-036', 'DST-060', 'DST-053', 'DST-069', 'DST-054', 'DST-063', 'DST-013', 'DST-040', 'DST-037', 'DST-070']	['DST-023', 'DST-021', 'DST-032', 'DST-054', 'DST-065']	['DST-054']	20%	5.88%	9.09%	14.60%
USR-00674	['DST-044', 'DST-041', 'DST-048', 'DST-036', 'DST-035', 'DST-049', 'DST-045', 'DST-070']	['DST-050', 'DST-023', 'DST-021', 'DST-041', 'DST-008']	['DST-041']	20%	12.5%	15.38%	13.12%

Table 12. Sample recommendation results generated by the hybrid

User Code	Ground Truth	Recommendation	Hits	Precision@5	Recall@5	F1@5	nDCG@5
USR-00031	['DST-032', 'DST-030', 'DST-025', 'DST-028', 'DST-036', 'DST-001', 'DST-065', 'DST-054', 'DST-037', 'DST-033', 'DST-042']	['DST-025', 'DST-048', 'DST-005', 'DST-054', 'DST-040']	['DST-025', 'DST-054']	40%	18.18%	25%	48.52%
USR-00032	['DST-014', 'DST-013', 'DST-056', 'DST-070', 'DST-036', 'DST-060', 'DST-053', 'DST-063', 'DST-066', 'DST-040', 'DST-004', 'DST-062', 'DST-041', 'DST-025', 'DST-055', 'DST-064', 'DST-069', 'DST-054', 'DST-037']	['DST-050', 'DST-041', 'DST-048', 'DST-053', 'DST-054']	['DST-041', 'DST-053', 'DST-054']	60%	15.78%	25%	65.48%
USR-00674	['DST-044', 'DST-041', 'DST-048', 'DST-064', 'DST-036', 'DST-035', 'DST-049', 'DST-045', 'DST-070']	['DST-050', 'DST-039', 'DST-041', 'DST-048', 'DST-040']	['DST-041', 'DST-048']	40%	22.22%	28.57%	27.72%

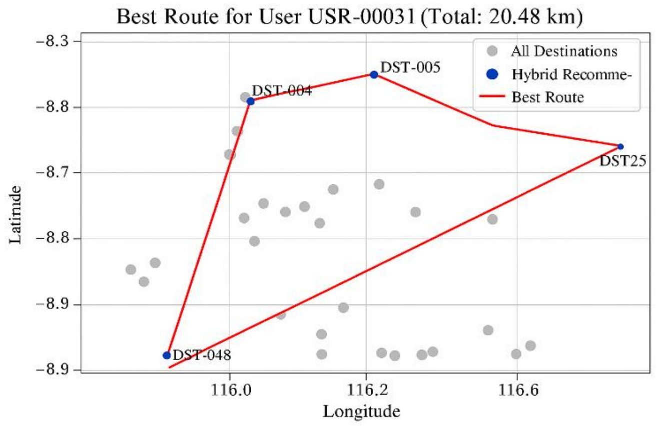

Figure 6. Optimized route for user USR-00031 using a genetic algorithm based on hybrid recommendations
Note: Top-k: ['DST-025', 'DST-048', 'DST-005', 'DST-054', 'DST-040'];
Route sequence: ['DST-005', 'DST-040', 'DST-048', 'DST-054', 'DST-025'].

Tables 10–12 present sample recommendation outputs for CBF, MF, and hybrid. The results demonstrate the distinct characteristics of each model. CBF generates items that closely match user preferences but lack diversity. MF captures alternative items beyond content features, but with low performance. Hybrid balances both approaches to provide more accurate and varied recommendations. An illustrative case can be seen in USR-00031, whose ground truth is dominated by the "*Pantai*" category (DST-036, DST-065, DST-054, DST-037, and DST-042), accompanied by several "*Wisata Alam*" (DST-032, DST-030, DST-025, DST-028, DST-033) and "*Air Terjun*" (DST-001). The CBF model strongly reflects this pattern by recommending mostly "*Pantai*" destinations (three out of five: DST-048, DST-054, DST-040) along with two "*Air Terjun*", showing high relevance but also overspecialization as the suggestions remain confined to similar categories. In contrast, MF introduces greater variety by incorporating *"Desa Wisata"* (DST-019 and DST-008) alongside *"Pantai"* (DST-050), *"Wisata Alam"* (DST-025), and *"Air Terjun"* (DST-003), although some recommendations are less well aligned with the ground truth.

Figure 6 illustrates the optimized travel route generated by the GA based on the hybrid top-k recommendations for user USR-00031. From the five selected destinations of the hybrid model (DST-025, DST-048, DST-005, DST-054, and DST-040), the algorithm arranges the visit sequence as DST-005 → DST-040 → DST-048 → DST-054 → DST-025 with a total estimated distance of 200.48 km. The red line connects the recommended destinations, while the blue points indicate the hybrid recommendation results. This optimization process not only provides a more efficient travel path but also assists users in planning their trips more effectively by considering travel distance while maintaining recommendation relevance. On average, the GA-optimized routes are 17.30% more efficient compared to the original hybrid routes. However, this improvement in route efficiency slightly reduces ranking quality, as indicated by the change in average nDCG@5 from 0.6237 (before GA) to 0.5814 (after GA).

4. CONCLUSION

This study presented a comprehensive evaluation of recommendation approaches for tourism applications, comparing CBF, MF, and a hybrid model combining CBF and MF. The results indicate that MF alone exhibits low precision and recall due to sparse user rating data, achieving only 15.08% precision, 7.11% recall, nDCG of 14.78%, and coverage of 40% under the 90% leave-out evaluation. RMSE visualizations confirm that rating predictions remain stable across different MF evaluation scenarios with varying leave-out proportions. In contrast, CBF improves performance by leveraging content similarity, reaching 48.69% precision, 21.64% recall, 52.58% nDCG, and 80% coverage, although it remains prone to overspecialization. The hybrid approach consistently outperforms both individual models, achieving the highest model performance with 59.16% precision, 25.98% recall, 60.95% nDCG, and 81.43% coverage, demonstrating the effectiveness of combining user interaction patterns with content-based features. Additionally, the application of a GA for route optimization provides users with a clear visitation sequence, facilitating travel planning and enhancing the usability, achieving an average route efficiency of 17.03%. However, this study is limited to a dataset of tourist destinations located exclusively on Lombok Island. For future research, it is recommended to explore context-aware and location-based hybrid models, incorporate real-time user feedback, and integrate multi-modal data such as images, reviews, and social media interactions to further enhance personalization and recommendation diversity. Moreover, extending the GA-based routing to consider travel constraints (e.g., time, cost, priority, and accessibility) can improve practical usability for itinerary planning.

ACKNOWLEDGMENT

We would like to express our gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, through the Directorate of Research, Technology, and Community Service, for providing funding for this research through the Fundamental Grant program for the 2025 fiscal year. Our appreciation also goes to the LPPM of Bumigora University for facilitating this research activity. Furthermore, we extend our thanks to the Communication and Informatics Office for their assistance in testing the level of technological readiness of this research.

REFERENCES

[1] Arici, H.E., Köseoglu, M.A. (2025). What are the most influential drivers of tourism destination competitiveness? Journal of Destination Marketing & Management, 36: 100990. https://doi.org/10.1016/j.jdmm.2025.100990
[2] Chen, X., Li, J. (2024). Facilitating knowledge-driven economic and social development: The significance of demographic transformation in tourism villages in China. Journal of the Knowledge Economy, 15(3): 13890-13918. https://doi.org/10.1007/s13132-023-01676-0
[3] Ridwan, M., Akther, A., Al Absy, M.S.M., et al. (2024). The role of tourism, technological innovation, and globalization in driving energy demand in major tourist regions. International Journal of Energy Economics and Policy, 14(6): 675-689. https://doi.org/10.32479/ijeep.17344
[4] Yang, C., Yang, J., Huang, J., You, B., Hong, H. (2024). An investigation of technological innovation and tourism

industrial structure upgrading as drivers of tourism economic growth. International Journal of Tourism Research, 26(4): e2687. https://doi.org/10.1002/jtr.2687
[5] Katyukha, D., Yurovchyk, V., Horiunova, K., Melnychenko, S., Myronov, Y. (2025). Leveraging cultural heritage for sustainable tourism growth: Contemporary strategies for conservation and promotion. OIDA International Journal of Sustainable Development, 18(7): 53-66. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5198993.
[6] Nguyen Thi, H., Nguyen Thi, T., Vu Trong, T., Nguyen Duc, T., Nguyen Nghi, T. (2024). Sustainable tourism governance: A study of the impact of culture. Journal of Governance and Regulation, 13(2): 474-485. https://doi.org/10.22495/jgrv13i2siart22
[7] Alam, M. (2025). Preserving cultural heritage and empowering indigenous communities for sustainable development in Fiji. Social Sciences & Humanities Open, 12: 101760. https://doi.org/10.1016/j.ssaho.2025.101760
[8] Pai, C.H., Zhang, Y., Wang, Y.L., Li, K., Shang, Y. (2025). Current challenges and opportunities in cultural heritage preservation through sustainable tourism practices. Current Issues in Tourism, 1-19. https://doi.org/10.1080/13683500.2024.2443776
[9] Erdős, F., Thinakaran, R., Firuza, B., Koloszár, L. (2025). The rise of AI in tourism-A systematic literature review. GeoJournal of Tourism & Geosites, 60. https://doi.org/10.30892/gtg.602spl22-1498
[10] Štilić, A., Puška, A., Nicić, M. (2025). The role of artificial intelligence in shaping the future of travel industry: An expert analysis of artificial intelligence-generated travel itineraries. DETUROPE-The Central European Journal of Regional Development and Tourism, 16(2): 57-79. https://doi.org/10.32725/det.2024.020
[11] Benbrahmi, F.Z., MEHOR, H.M.H. (2025). The impact of artificial intelligence on enhancing the travel experience for tourists: An analysis of successful models. Business Excellence and Management, 15(5): 5-13. https://doi.org/10.24818/beman/2025.S.I.5-01
[12] Florido-Benítez, L., del Alcázar Martínez, B. (2024). How artificial intelligence (AI) is powering new tourism marketing and the future agenda for smart tourist destinations. Electronics, 13(21): 4151. https://doi.org/10.3390/electronics13214151
[13] Solano-Barliza, A., Arregocés-Julio, I., Aarón-Gonzalvez, M., et al. (2024). Recommender systems applied to the tourism industry: A literature review. Cogent Business & Management, 11(1): 2367088. https://doi.org/10.1080/23311975.2024.2367088
[14] Sharma, R. (2024). Genetic algorithm based personalized travel recommendation system. In 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, pp. 867-874. https://doi.org/10.1109/IDCIoT59759.2024.10467470
[15] Alsahafi, R., Mehmood, R., Alqahtany, S. (2025). A machine learning-based analysis of tourism recommendation systems: Holistic parameter discovery and insights. International Journal of Advanced Computer Science & Applications, 16(1): 1369-1382. https://doi.org/10.14569/ijacsa.2025.01601130
[16] Nguyen, L.V. (2024). OurSCARA: Awareness-based recommendation services for sustainable tourism. World, 5(2): 471-482. https://doi.org/10.3390/world5020024
[17] Chrysafiadi, K., Kontogianni, A., Virvou, M., Alepis, E. (2025). Enhancing user experience in smart tourism via fuzzy logic-based personalization. Mathematics, 13(5): 846. https://doi.org/10.3390/math13050846
[18] Alfaifi, Y.H. (2024). Recommender systems applications: Data sources, features, and challenges. Information, 15(10): 660. https://doi.org/10.3390/info15100660
[19] Ariyani, N., Fauzi, A. (2024). Unlocking sustainable rural tourism to support rural development: A Bayesian approach to managing water-based destinations in Indonesia. Sustainability, 16(13): 5506. https://doi.org/10.3390/su16135506
[20] Hammad, R., Azwar, M., Syarif, M.A. (2025). Optimizing tourism recommendations with a hybrid model: Bridging user preferences and behavioral patterns. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 9(4): 894-904. https://doi.org/10.29207/resti.v9i4.6510
[21] Lestari, Y.B., Yusra, K. (2022). Identifying tourism potentials of ethno-cultural attractions in Lombok. Sustainability, 14(23): 16075. https://doi.org/10.3390/su142316075
[22] Hammad, R., Irfan, P., Mukti, M.T.P. (2024). Geolocation data incorporation in Mapbox for comprehensive mapping of tourism areas on Lombok Island. Matrix: Jurnal Manajemen Teknologi dan Informatika, 14(1): 33-42. http://doi.org/10.31940/matrix.v14i1.33-42
[23] Praditya, N.W.P.Y., Permanasari, A.E., Hidayah, I. (2021). Designing a tourism recommendation system using a hybrid method (collaborative filtering and content-based filtering). In 10th IEEE International Conference on Communication, Networks and Satellite, Purwokerto, Indonesia, pp. 298-305. https://doi.org/10.1109/COMNETSAT53002.2021.9530823
[24] Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., Sabri, A. (2021). Hybrid recommender system for tourism based on big data and AI: A conceptual framework. Big Data Mining and Analytics, 4(1): 47-55. https://doi.org/10.26599/BDMA.2020.9020015
[25] Garipelly, V., Adusumalli, P.T., Singh, P. (2021). Travel recommendation system using content and collaborative filtering-a hybrid approach. In 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-4. https://doi.org/10.1109/ICCCNT51525.2021.9579907
[26] Forouzandeh, S., Rostami, M., Berahmand, K. (2022). A hybrid method for recommendation systems based on tourism with an evolutionary algorithm and TOPSIS model. Fuzzy Information and Engineering, 14(1): 26-50. https://doi.org/10.1080/16168658.2021.2019430
[27] Aldayel, M., Al-Nafjan, A., Al-Nuwaiser, W.M., Alrehaili, G., Alyahya, G. (2023). Collaborative filtering-based recommendation systems for touristic businesses, attractions, and destinations. Electronics, 12(19): 4047. https://doi.org/10.3390/electronics12194047
[28] Chang, J.L., Li, H., Bi, J.W. (2022). Personalized travel recommendation: A hybrid method with collaborative filtering and social network analysis. Current Issues in Tourism, 25(14): 2338-2356.

https://doi.org/10.1080/13683500.2021.2014792
[29] Regula, T., Ali, A. (2024). Enhancing tourism recommendation system using KNN-based hybrid mobile application. In International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, pp. 1591-1594. https://doi.org/10.1109/ICSCNA63714.2024.10863856
[30] Wang, Z. (2023). Intelligent recommendation model of tourist places based on collaborative filtering and user preferences. Applied Artificial Intelligence, 37(1): 2203574. https://doi.org/10.1080/08839514.2023.2203574
[31] Muneer, M., Rasheed, U., Khalid, S., Ahmad, M. (2022). Tour spot recommendation system via content-based filtering. In 2022 16th International Conference on Open Source Systems and Technologies (ICOSST), Lahore, Pakistan, pp. 1-6. https://doi.org/10.1109/ICOSST57195.2022.10016820
[32] Huda, C., Heryadi, Y., Budiharto, W. (2024). Smart tourism recommender system modeling based on hybrid technique and content boosted collaborative filtering. IEEE Access, 12: 131794-131808. https://doi.org/10.1109/ACCESS.2024.3450882
[33] Yan, J., Wang, M., Xue, R. (2025). Hybrid recommendation system based on human-computer interaction technology and smart tourism. International Journal of Ad Hoc and Ubiquitous Computing, 48(1): 1-12. https://doi.org/10.1504/IJAHUC.2025.143545
[34] Chen, Z., Zhang, P., Peng, L. (2024). Application of a hybrid genetic algorithm based on the travelling salesman problem in rural tourism route planning. International Journal of Computing Science and Mathematics, 19(1): 1-14. https://doi.org/10.1504/IJCSM.2024.136816
[35] Istiqfarri, A.Q., Baizal, Z.K.A., Wulandari, G.S. (2025). Hybrid genetic and simulated annealing algorithm for vehicle routing problem in recommendation of tour route in days. International Journal of Intelligent Engineering & Systems, 18(3): 257-272. https://doi.org/10.22266/ijies2025.0430.18
[36] Lakshmi, J.V.N., Pallavi, M.O. (2024). A hybrid recommendation model for tourist using evolutionary algorithm combined with local search algorithm for trip planning. SN Computer Science, 5(6): 743. https://doi.org/10.1007/s42979-024-03063-1
[37] Kumar, V., Chin, Y.B., Soong, L.W. (2024). Intelligent tourist attractions recommender system with hybrid collaborative filtering. In 2024 Multimedia University Engineering Conference (MECON), Cyberjaya, Malaysia, pp. 1-4. https://doi.org/10.1109/MECON62796.2024.10776473
[38] Zhang, S. (2025). Integrating user profiles and collaborative filtering for smart recommendation of tourism city cultural and creative products. International Journal of High Speed Electronics and Systems, 34(4): 2540296. https://doi.org/10.1142/S0129156425402967
[39] Yuan, Y., Zheng, W. (2024). Your trip, your way: An adaptive tourism recommendation system. Applied Soft Computing, 154, 111330. https://doi.org/10.1016/j.asoc.2024.111330
[40] Bolotbekova, A., Hakli, H., Beskirli, A. (2025). Trip route optimization based on bus transit using genetic algorithm with different crossover techniques: A case study in Konya/Türkiye. Scientific Reports, 15(1): 2491. https://doi.org/10.1038/s41598-025-86695-4
[41] Siino, M., Tinnirello, I., La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers. Information Systems, 121, 102342. https://doi.org/10.1016/j.is.2023.102342
[42] Sulistianingsih, N., Switrayana, I.N. (2024). Enhancing sentiment analysis for the 2024 Indonesia election using SMOTE-Tomek Links and binary logistic regression. International Journal of Education and Management Engineering, 14(3): 22-32. https://doi.org/10.5815/ijeme.2024.03.03
[43] Mohammadi, M., Naree, S.A., Lati, M. (2020). User-item content awareness in matrix factorization based collaborative recommender systems. Intelligent Data Analysis, 24(3): 723-739. https://doi.org/10.3233/IDA-194599
[44] Singh, R.H., Maurya, S., Tripathi, T., Narula, T., Srivastav, G. (2020). Movie recommendation system using cosine similarity and KNN. International Journal of Engineering and Advanced Technology, 9(5): 556-559. https://doi.org/10.35940/ijeat.e9666.069520
[45] Abdurrafi, M.F., Ningsih, D.H.U. (2023). Content-based filtering using cosine similarity algorithm for alternative selection on training programs. Journal of Soft Computing Exploration, 4(4): 204-212. https://doi.org/10.52465/joscex.v4i4.232
[46] Zhang, Y. (2022). An introduction to matrix factorization and factorization machines in recommendation system, and beyond. arXiv preprint arXiv:2203.11026. https://doi.org/10.48550/arXiv.2203.11026
[47] Alabduljabbar, R. (2023). Matrix factorization collaborative-based recommender system for Riyadh restaurants: Leveraging machine learning to enhance consumer choice. Applied Sciences, 13(17): 9574. https://doi.org/10.3390/app13179574
[48] Bobadilla, J., Dueñas Lerín, J., Ortega, F., Gutierrez, A. (2024). Comprehensive evaluation of matrix factorization models for collaborative filtering recommender systems. International Journal of Interactive Multimedia and Artificial Intelligence, 8(6): 15-23. https://doi.org/10.9781/ijimai.2023.04.008
[49] Switrayana, I.N., Maulidevi, N.U. (2022). Collaborative convolutional autoencoder for scientific article recommendation. In 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, pp. 96-101. https://doi.org/10.1109/ICITACEE55701.2022.9924130
[50] Muhammad, M., Gunardi, G., Danardono, D., Rosadi, D. (2025). Optimization hybrid weighted switching filtering (OHWSF) using SVD and SVD++ for addressing data sparsity. International Journal of Advances in Intelligent Informatics, 11(3): 533-549. https://doi.org/10.26555/ijain.v11i3.1796
[51] Beloborodov, D., Zimovnov, A., Molodyk, P., Kirillov, D. (2022). Efficient mixed dimension embeddings for matrix factorization. arXiv preprint arXiv:2205.11248. https://doi.org/10.48550/arXiv.2205.11248
[52] Costa, A., Roda, F. (2011). Recommender systems by means of information retrieval. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, Sogndal, Norway, pp. 1-5.

https://doi.org/10.1145/1988688.1988755
[53] Oh, S., Ustun, B., McAuley, J., Kumar, S. (2022). Rank list sensitivity of recommender systems to interaction perturbations. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA pp. 1584-1594. https://doi.org/10.1145/3511808.3557425
[54] Saini, K., Singh, A. (2024). Original research article comparative analysis of collaborative filtering recommender system algorithms for e-commerce. Journal of Autonomous Intelligence, 7(2): 1-12. https://doi.org/10.32629/jai.v7i2.1182
[55] Choi, K.C., Li, S., Lam, C.T., et al. (2022). Genetic algorithm for tourism route planning considering time constrains. International Journal of Engineering Trends and Technology, 70(1): 170-178. https://doi.org/10.14445/22315381/IJETT-V70I1P219
[56] Switrayana, I.N., Osmond, A.B., Aditsania, A. (2019). Item delivery simulation using genetic algorithm. Journal of Physics: Conference Series, 1201(1): 012060. https://doi.org/10.1088/1742-6596/1201/1/012060
[57] Rahman, I., Mathieson, L., Ahamed, F. (2024). Comparative analysis of genetic algorithm, ant colony optimisation and particle swarm optimisation on the travelling salesman problem and the 0/1 knapsack problem. In 2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS), Sydney, Australia, pp. 429-434. https://doi.org/10.1109/FMLDS63805.2024.00082
[58] Li, H. (2025). Solving the TSP problem based on improved genetic algorithm. Applied and Computational Engineering, 133: 73-79. https://doi.org/10.54254/2755-2721/2025.20603
[59] Hussain, A., Riaz, S., Amjad, M.S., Haq, E.U. (2022). Genetic algorithm with a new round-robin based tournament selection: Statistical properties analysis. PloS One, 17(9): e0274456. https://doi.org/10.1371/journal.pone.0274456
[60] Alfarizi, M.G., Stanko, M., Bikmukhametov, T. (2022). Well control optimization in waterflooding using genetic algorithm coupled with Artificial Neural Networks. Upstream Oil and Gas Technology, 9, 100071. https://doi.org/10.1016/j.upstre.2022.100071
[61] Jadon, A., Patil, A. (2025). A comprehensive survey of evaluation techniques for recommendation systems. Computation of Artificial Intelligence & Machine Learning, 2185: 281-304. https://doi.org/10.1007/978-3-031-71484-9_25
[62] Gunathilaka, T.M.A.U., Manage, P.D., Zhang, J., Li, Y., Kelly, W. (2025). Addressing sparse data challenges in recommendation systems: A systematic review of rating estimation using sparse rating data and profile enrichment techniques. Intelligent Systems with Applications, 25: 200474. https://doi.org/10.1016/j.iswa.2024.200474
[63] Sabiri, B., Khtira, A., El Asri, B., Rhanoui, M. (2025). Hybrid quality-based recommender systems: A systematic literature review. Journal of Imaging, 11(1): 12. https://doi.org/10.3390/jimaging11010012
[64] Raza, S., Rahman, M., Kamawal, S., et al. (2024). A comprehensive review of recommender systems: Transitioning from theory to practice. arXiv preprint arXiv:2407.13699. https://doi.org/10.48550/arXiv.2407.13699