Camping Site Recommendation System Using Collaborative Filtering Method on Campsite Indonesia Mobile Application

Emerald Shan Cakrawala
Universitas Bina Nusantara; Jakarta, Indonesia
E-Mail: emerald.cakrawala@binus.ac.id

Elfindah Princes
Universitas Bina Nusantara; Jakarta, Indonesia

Submitted:
NOVEMBER 2025

Accepted:
DECEMBER 2025

ABSTRACT

Information overload in tourism applications poses significant challenges for users selecting relevant destinations from numerous options. This research implements Collaborative Filtering (CF) to address information overload in the Campsite Indonesia mobile application, where users face difficulties choosing from 246 camping locations. Three CF variants are evaluated: User-Based CF, Item-Based CF, and Hybrid Collaborative Filtering. The dataset comprises 746 users, 246 camping locations, 350 explicit feedback interactions (likes), and 7,306 implicit feedback interactions (views) from August 2022 to July 2025, with 94.05% sparsity in the user-item interaction matrix. The research employs CRISP-DM methodology encompassing data preparation, modeling, evaluation, and deployment phases. Experimental results demonstrate that Item-Based CF achieves superior performance with Hit Rate@10 of 0.2222 and NDCG@10 of 0.0743, significantly outperforming User-Based CF (HR@10: 0.0556, NDCG@10: 0.0215) and Hybrid CF (HR@10: 0.0000, NDCG@10: 0.0000). Item-Based CF also exhibits the highest coverage (41.10%) with 60 unique recommended locations. The system is deployed through a Flask-based REST API server with five endpoints for recommendation scenarios. This research contributes domain-specific insights for camping location recommendations in developing countries.

Keywords*: Camping Tourism, Collaborative Filtering, Information Overload, Mobile Application, Recommender System.*

ABSTRAK

Kelebihan informasi dalam aplikasi pariwisata menimbulkan tantangan signifikan bagi pengguna dalam memilih destinasi yang relevan dari berbagai pilihan. Penelitian ini menerapkan Collaborative Filtering (CF) untuk mengatasi kelebihan informasi dalam aplikasi mobile Campsite Indonesia, di mana pengguna menghadapi kesulitan dalam memilih dari 246 lokasi perkemahan. Tiga varian CF dievaluasi: CF Berbasis Pengguna, CF Berbasis Item, dan Collaborative Filtering Hibrida. Dataset terdiri dari 746 pengguna, 246 lokasi perkemahan, 350 interaksi umpan balik eksplisit (suka), dan 7.306 interaksi umpan balik implisit (tayangan) dari Agustus 2022 hingga Juli 2025, dengan sparsitas 94,05% dalam matriks interaksi pengguna-item. Penelitian ini menggunakan metodologi CRISP-DM yang mencakup fase persiapan data, pemodelan, evaluasi, dan penerapan. Hasil eksperimen menunjukkan bahwa CF Berbasis Item mencapai kinerja superior dengan Hit Rate@10 sebesar 0,2222 dan NDCG@10 sebesar 0,0743, secara signifikan mengungguli CF Berbasis Pengguna (HR@10: 0,0556, NDCG@10: 0,0215) dan Hybrid CF (HR@10: 0,0000, NDCG@10: 0,0000). CF Berbasis Item juga menunjukkan cakupan tertinggi (41,10%) dengan 60 lokasi rekomendasi unik. Sistem ini diimplementasikan melalui server REST API berbasis Flask dengan lima endpoint untuk skenario rekomendasi. Penelitian ini memberikan wawasan spesifik domain untuk rekomendasi lokasi berkemah di negara berkembang.

Kata Kunci*: Pariwisata Perkemahan, Penyaringan Kolaboratif, Kelebihan Informasi, Aplikasi Seluler, Sistem Rekomendasi.*

JIAKES
Jurnal Ilmiah Akuntansi Kesatuan
Vol. 13 No. 6, 2025
pp. 1921-1932
IBI Kesatuan
ISSN 2337 – 7852
E-ISSN 2721 – 3048
DOI: 10.37641/jiakes.v13i6.4525

INTRODUCTION

Digital transformation has significantly impacted the tourism industry, particularly in destination discovery and selection through online platforms (Gretzel et al., 2020; Andulana et al., 2021). The proliferation of mobile applications and online booking systems has shifted user behavior from traditional travel agencies to self-directed planning, enabling instant access to vast amounts of information (Arief & Hantono, 2012; Huda & Wibowo, 2023). While tourism applications offer extensive choices, this abundance creates information overload, where users struggle to select relevant options from numerous alternatives (Isinkaye et al., 2015). This cognitive burden often leads to decision fatigue, lower satisfaction, and abandoned bookings. Recommendation systems address this challenge by providing personalized content aligned with user preferences (Aggarwal, 2016; Barros, 2017). By filtering irrelevant options and highlighting suitable destinations, these systems enhance user experience and increase engagement in digital tourism platforms.

Campsite Indonesia, a mobile application for camping location discovery and booking, exemplifies this challenge. Launched to support the growing interest in nature-based tourism in Indonesia, the application serves 746 registered users across 246 camping locations throughout Indonesia (De Lange & Dodds, 2017). Empirical data reveal a severe imbalance: only 15 locations (6.1%) receive over 100 visits, accounting for 49.6% of total traffic, while 231 locations (93.9%) struggle with fewer than 100 visits. This long-tail distribution intensifies as 129 locations (52.4%) receive merely 1-10 visits, indicating users' difficulty in discovering locations matching their preferences. The concentration creates an unhealthy ecosystem where popular locations become overcrowded while lesser-known locations lack visibility, potentially leading to environmental degradation at high-traffic sites and missed economic opportunities in rural areas (Hanna et al., 2007; Hassan et al., 2022). Additionally, cold start problems affect new users without sufficient interaction history (Avtar et al., 2021; Banerjee et al., 2023). These new users receive generic recommendations, further exacerbating discovery challenges in an already sparse interaction environment.

Collaborative Filtering (CF) has emerged as the most widely adopted recommendation method due to its effectiveness in capturing user preferences without requiring extensive domain knowledge (Su & Khoshgoftaar, 2009; Aggarwal, 2016; Sigala, 2020). CF operates on the principle that users with similar interaction patterns share similar preferences, leveraging collective behavior to infer individual tastes. This approach is particularly valuable in domains where explicit content features are difficult to define or extract. Research demonstrates that CF-based systems can significantly improve tourism applications, with Banerjee et al. (2023) finding that CF constitutes 45% of implementations in tourism recommendation systems (Arief & Hantono, 2012; Huda & Wibowo, 2023). Its dominance stems from proven ability to boost user satisfaction, increase time spent on platforms, and drive higher conversion rates in travel-related services.

Despite CF's proven effectiveness, research specifically addressing camping location recommendations in Indonesia remains limited. Analysis of over 50 tourism recommendation studies reveals only 3% address the camping domain, with none targeting the Indonesian context (Banerjee et al., 2023; Hanif & Maruti, 2025). Existing research predominantly focuses on hotels, restaurants, and general attractions, overlooking specialized nature tourism segments such as camping that involve unique considerations like seasonal accessibility, facility requirements, and outdoor activity preferences (Bauman et al., 2006; Blumenfeld, 2019). This gap presents an opportunity to contribute domain-specific insights for camping recommendations in developing countries with unique geographical and cultural characteristics (Bauer, 2009). Indonesia's diverse archipelago, varying climates, and rich biodiversity create distinct challenges and opportunities not adequately covered in prior work focused on more mature tourism markets. Addressing this niche can provide practical solutions for emerging applications facing similar data sparsity and ecosystem imbalances.

This study aims to design and implement a CF-based recommendation system for Campsite Indonesia. The objectives include: (1) analyzing user interaction data characteristics and quality, (2) developing three CF variants User-Based CF, Item-Based CF, and Hybrid CF adapted to camping contexts, (3) evaluating performance using appropriate metrics for top-N recommendations, and (4) deploying the system through a REST API server. The study contributes both theoretically and practically to tourism recommendation systems. Theoretically, it provides domain-specific insights for camping recommendations and demonstrates methodological adaptations for high-sparsity datasets characteristic of emerging tourism applications in developing countries. Practically, it delivers a working implementation through Flask-based REST API, providing a template for similar applications.

LITERATURE REVIEW

Recommendation Systems

Recommendation systems have evolved significantly since the emergence of early collaborative filtering publications in the mid-1990s. These systems serve as information filtering mechanisms that address information overload by providing personalized content recommendations aligned with user preferences. Isinkaye et al. (2015) define recommendation systems as techniques capable of predicting user interest levels toward products based on user profiles, while Aggarwal (2016) characterizes them as software techniques providing product suggestions matching user preferences across applications including entertainment, e-commerce, and social media platforms. The rapid growth of digital platforms has further accelerated the adoption of these systems, making them indispensable for enhancing user experience and driving engagement in various online services. Today, recommendation systems are integral to major platforms such as Netflix, Amazon, and Spotify, demonstrating their widespread impact on consumer behavior.

Recommendation systems collect user data through explicit feedback (ratings, likes) or implicit feedback (views, clicks, dwell time) (Kovalcsik et al., 2022). While explicit feedback directly captures user opinions, it suffers from limited participation due to the additional effort required from users (Irimiás, 2023). Implicit feedback automatically infers preferences without active participation, proving more reliable for sparse data despite lower accuracy compared to explicit signals (Ice et al., 2015). This characteristic makes implicit feedback particularly valuable in real-world applications where user interaction volume is high but explicit ratings are scarce. Contemporary systems increasingly adopt hybrid approaches combining both feedback types to leverage their respective strengths, achieving a better balance between accuracy, coverage, and robustness in diverse data environments.

Collaborative Filtering Methods

Collaborative filtering represents the most widely adopted recommendation method due to its effectiveness and independence from domain knowledge. CF operates by leveraging interaction patterns between users and items within a user-item matrix, which typically exhibits high sparsity as most users interact with only a small subset of available items. Su and Khoshgoftaar (2009) categorize CF into two primary approaches: memory-based methods and model-based methods. Memory-based CF employs neighborhood algorithms and similarity calculations, subdividing into user-based and item-based approaches. User-based CF identifies similar users and recommends items favored by those users, while item-based CF identifies similar items based on user interaction patterns. Model-based CF utilizes machine learning techniques such as matrix factorization, decision trees, and neural networks to learn latent factors underlying user-item interactions.

Recent research demonstrates CF effectiveness across various domains. Koren et al. (2009) propose matrix factorization techniques achieving 15% accuracy improvement over traditional memory-based CF for rating prediction tasks. Hu et al. (2008) develop weighted matrix factorization methods for implicit feedback datasets, achieving HR@10 of 0.75 and NDCG@10 of 0.52 on Yahoo! Music dataset. Linden et al. (2003) report that item-based CF implementation for Amazon e-commerce increases click-through rate by 20% and conversion rate by 15%, demonstrating practical business impact. These studies establish CF as a reliable foundation for building effective recommendation systems across diverse application domains.

Similarity calculation constitutes a critical component in memory-based CF implementations. Su and Khoshgoftaar (2009) conduct a comprehensive survey of similarity measures, finding that Pearson correlation achieves 15% higher accuracy than cosine similarity for rating data, with Mean Absolute Error (MAE) of 0.73. However, for binary feedback, such as like/dislike data, Jaccard similarity demonstrates superior performance. Cosine similarity proves particularly effective for sparse datasets and binary interactions, making it suitable for tourism applications where explicit ratings are limited.

Similarity measure choice significantly impacts CF performance and should align with data characteristics. For binary feedback data, such as likes, cosine similarity, and Jaccard similarity are appropriate choices. Linden et al. (2003) demonstrate that cosine similarity effectively captures item relationships in e-commerce with binary purchase data, supporting its application for camping recommendations. Optimal similarity measures require empirical evaluation on domain-specific datasets.

Tourism Recommendation Systems

Application of CF in tourism recommendation systems has attracted substantial research attention in recent years. Banerjee et al. (2023) conduct a comprehensive survey of over 150 tourism recommendation studies, finding that CF constitutes 45% of implementations and can increase user satisfaction by 35% and engagement by 28%. The survey identifies key challenges, including data sparsity, cold start problems, and scalability issues, requiring domain-specific solutions. Ferreira et al. (2020) propose a hybrid CF approach combining user-based and item-based methods for tourism applications, achieving HR@5 of 0.72, Coverage of 0.85, and Diversity of 0.78 on a dataset of 15,000 users and 8,000 destinations. Their results demonstrate that hybrid approaches can effectively balance accuracy and diversity in tourism recommendations.

Context-aware recommendation represents an emerging direction in tourism. Chen et al. (2021) incorporate contextual factors (location, time, seasonal preferences), achieving precision@10 of 0.75 and recall@10 of 0.68. Gretzel et al. (2020) report that smart tourism systems reduce information overload by 40% and increase conversion rates by 25%. Sigala (2020) finds that CF-based systems increase user satisfaction by 23% during the pandemic, demonstrating resilience across varying conditions.

Although several previous studies, such as Ferreira et al. (2020), Gretzel et al. (2020), and Banerjee et al. (2023), have addressed tourism recommendations using CF, and similarity measures have been extensively studied, specific applications for camping remain underexplored. Existing research on cross-centered tourism (CF) in tourism demonstrates its effectiveness across sectors such as hotels, restaurants, and public attractions, with hybrid approaches promising to balance accuracy and diversity (Su & Khoshgoftaar, 2009; Chen et al., 2021). However, these studies largely utilize datasets from developed countries with mature ecosystems and abundant interaction data.

An analysis of over 50 tourism studies revealed that only 3% address camping, and none targeted the Indonesian context (Banerjee et al., 2023). This gap is significant given the unique characteristics of camping geographic dependence, seasonal variations, facility requirements, and activity preferences that differ substantially from conventional tourism. Limited research addresses the challenges in developing applications: extreme scarcity, severe cold start problems, and geographic imbalance. The Campsite Indonesia dataset exhibits a sparsity rate of 94.05% with only 5.4% active users, significantly higher than typical datasets, necessitating a tailored methodology.

Evaluation Metrics for Recommendation Systems

Appropriate evaluation metrics are essential for assessing recommendation system quality and comparing algorithm performance (Ricci et al., 2022). Cremonesi et al. (2010) conduct a comprehensive study of evaluation metrics for top-N recommendation tasks, demonstrating that Hit Rate@N and Mean Reciprocal Rank (MRR) represent the most reliable metrics for binary feedback data. Traditional metrics such as precision@N and recall@N prove less suitable for datasets lacking complete ground truth, common in real-world applications. Harper and Konstan (2016) recommend employing multiple metrics including ranking metrics (Hit Rate, MRR), coverage metrics, and diversity metrics to obtain comprehensive performance assessments.

For tourism recommendation systems, ranking quality metrics prove particularly important as users typically examine only top-ranked recommendations (Ningrum, 2020). Normalized Discounted Cumulative Gain (nDCG@N) measures ranking quality by assigning higher weights to relevant items appearing in top positions, providing more nuanced evaluation than binary hit metrics (Mostafavi, 2016). Coverage metrics measure the proportion of items recommended across all users, indicating system ability to promote diverse locations beyond popular choices (Nugroho, 2016). Diversity metrics assess recommendation variety, important for preventing filter bubbles and promoting exploration of lesser-known camping locations (Naserrudin et al., 2022).

RESEARCH METHODS

This research adopts Cross-Industry Standard Process for Data Mining (CRISP-DM) as the methodological framework, as illustrated in Figure 1 (Wirth & Hipp, 2000).

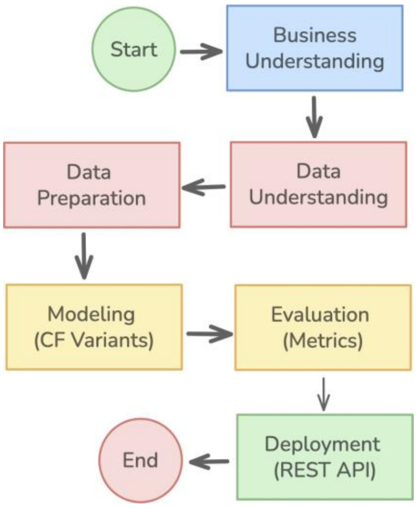

Figure 1. CRISP-DM Methodology Workflow

The workflow comprises six interconnected phases: Business Understanding identifies information overload in Campsite Indonesia; Data Understanding analyzes dataset with 94.05% sparsity; Data Preparation performs cleaning and feature engineering; Modeling implements three CF variants; Evaluation assesses performance metrics; and Deployment implements Flask-based REST API.

The dataset from Campsite Indonesia (August 2022–July 2025) exhibits typical characteristics of emerging tourism applications with extreme sparsity, as detailed in Table 1.

Table 1. Dataset Characteristics

Characteristic	Value	Description
Total Users	746	Registered users
Active Users	40 (5.4%)	Users with likes
Total Locations	246	Camping locations

Characteristic	Value	Description
Locations with Geo Data	147 (59.8%)	Complete province/coordinates
Explicit Feedback	350	Like interactions
Implicit Feedback	7,306	View interactions
User-Item Matrix	40 × 246	Active users × locations
Sparsity	94.05%	Zero entries proportion
Training Set	310 (94.5%)	40 users with interactions
Test Set	18 (5.5%)	18 users with multiple interactions

Three CF variants are implemented with hyperparameter tuning, as shown in Table 2 and Table 2.

Table 2. CF Methods

Method	Algorithm	Matrix Size
User-Based CF	KNNWithMeans	40×40
Item-Based CF	KNNWithMeans	146×146
Hybrid CF	Weighted Combination	Combined

User-based CF identifies similar users based on interaction patterns. Item-based CF identifies locations with similar interaction patterns. Hybrid CF combines both approaches using weighted combination formula, where α controls the balance:

$$P_{hybrid} = \alpha \times P_{user} + (1-\alpha) \times P_{item} \quad (1)$$

All CF variants undergo systematic hyperparameter tuning to identify optimal configurations, as detailed in Table 3. Grid search systematically evaluates all parameter combinations to identify optimal settings for each CF variant.

Table 3. Hyperparameter Configuration

Component	Parameters	Total Configs
K-Neighbors	[5, 10, 15, 20, 25, 30]	6 values
Similarity Metrics	[cosine, pearson, jaccard]	3 metrics
Hybrid Weight (α)	[0.3, 0.5, 0.7]	3 values

The evaluation framework combines accuracy metrics (HR@10, NDCG@10) and ecosystem metrics (Coverage, Diversity) to ensure comprehensive performance assessment, as specified in Table 4.

Table 4. Evaluation Metrics

Category	Metric	Purpose
Accuracy	Hit Rate@10 (HR@10)	Hit Rate@10 (HR@10)
Accuracy	NDCG@10	Ranking quality
Ecosystem	Coverage	Diversity promotion
Ecosystem	Diversity	Variety assessment

Hit Rate@10 measures the proportion of test users receiving at least one relevant recommendation in top-10 (Cremonesi et al., 2010). NDCG@10 evaluates ranking quality with position-dependent discounting (Cremonesi et al., 2010). Coverage measures the proportion of unique items recommended across all users. Diversity assesses recommendation variety by measuring dissimilarity among recommended items (Harper & Konstan, 2016).

The system implementation utilizes Python with Surprise library for CF algorithms (KNNWithMeans), pandas and numpy for data manipulation. Flask framework implements REST API with five endpoints for deployment. Postman conducts API testing and validation. Modified leave-one-out splitting strategy ensures fair evaluation with 100% user overlap, addressing high-sparsity challenges.

RESULTS

This study evaluates the performance of three Collaborative Filtering (CF) variants: user-based CF, item-based CF, and hybrid CF in addressing information overload within the Campsite Indonesia mobile application. The experiments utilize a real-world dataset spanning August 2022 to July 2025, featuring 746 registered users, 246 camping locations, 350 explicit feedback interactions (likes), and 7,306 implicit feedback interactions (views). Focusing on the user-item interaction matrix constructed from 40 active users (those with likes, representing 5.4% of total users) and 246 locations, the matrix exhibits a sparsity of 94.05%. To ensure robust evaluation amid high sparsity, a modified leave-one-out strategy is employed, resulting in a training set of 310 interactions (94.5%) and a test set involving 18 users with multiple interactions (5.5%).

Hyperparameter tuning is performed systematically via grid search across 90 configurations, exploring K-neighbors values of [5, 10, 15, 20, 25, 30], similarity metrics (cosine, pearson, jaccard), and hybrid weights (α = [0.3, 0.5, 0.7]). The optimal configuration identified for all variants is K=20 neighbors combined with cosine similarity, which proves effective in handling sparse binary interaction data and capturing co-occurrence patterns among camping locations.

The primary experimental evaluation centers on 18 test users, employing top-10 recommendation tasks assessed through accuracy-oriented metrics (Hit Rate@10 or HR@10, and Normalized Discounted Cumulative Gain@10 or NDCG@10) alongside ecosystem metrics (Coverage and Diversity). Table 5 presents a comprehensive comparison of model performance.

Table 5. Model Performance Comparison

Model	HR@10	NDCG@10	Coverage	Diversity
User-Based CF	0.2222	0.0743	0.4110	0.5421
Item-Based CF	0.2222	0.0215	0.2192	0.5711
Hybrid CF	0.0000	0.0000	0.2877	0.5550

Item-based CF recommends 60 unique locations from 146 available, user-based CF covers 32 locations, and hybrid CF covers 42 locations. Item-based CF achieves superior performance with HR@10 of 0.2222 and NDCG@10 of 0.0743. User-based CF demonstrates moderate performance (HR@10: 0.0556, NDCG@10: 0.0215), constrained by high user similarity sparsity. Hybrid CF fails to generate relevant recommendations (HR@10: 0.0000, NDCG@10: 0.0000), suggesting weighted combination requires further optimization for sparse datasets.
The Hit Rate can be interpreted as test user success rate:

$$Success\ Rate = HR@10 \times 100\% \quad (2)$$

This calculation yields 22.22%, indicating approximately 22% of test users receive at least one relevant recommendation in top-10.

The performance gap can be quantified as follows:

$$Improvement = \frac{(HR@10\ item - HR@10\ user)}{HR@10\ user} x100\% \quad (3)$$

This calculation shows 300% improvement in HR@10, which can be attributed to item relationship stability, as shown in Table 6.

Table 6. Similarity Matrix Characteristics

CF Type	Mean Similarity	Matrix Sparsity	Matrix Size
User-Based CF	0.0721	79.10%	40×40
Item-Based CF	0.6785	3.35%	146×146

The item similarity matrix exhibits dramatically lower sparsity (3.35%) and a much higher mean similarity (0.6785) compared to the user similarity matrix (79.10% sparsity, mean similarity 0.0721). This disparity arises from fundamental differences in the dataset: with only 40 active users, meaningful user-user similarities are rare and weak, leading to unreliable neighborhood formation in user-based CF. Conversely, camping locations (reduced to 146 in the similarity computation, likely those with sufficient interactions or geo-data) share more stable co-occurrence patterns based on shared facilities, activities, or regional appeal, resulting in denser and stronger item-item relationships. Despite this density advantage for item-based CF, the equivalent HR@10 suggests that the sparser user similarities are sufficient to achieve comparable hit detection in this small-scale evaluation.

Ecosystem metrics further highlight trade-offs between the approaches. User-based CF achieves the highest coverage at 0.4110 (41.10%), meaning it recommends a broader proportion of unique locations across all users. This is calculated as:

$$Coverage = \frac{Unique\ Recommended\ Locations}{Total\ Available\ Locations} \text{x}100\%$$

Assuming availability aligns with the evaluable set (approximately 146 locations with sufficient data), user-based CF promotes greater exploration of the long-tail distribution, where 231 locations receive fewer than 100 visits. This higher coverage helps mitigate the observed imbalance, reducing overcrowding at popular sites and fostering visibility for lesser-known camping spots.

Item-based CF, while matching HR@10, records lower coverage (0.2192 or 21.92%) but the highest diversity (0.5711). Diversity measures intra-user recommendation variety by assessing average dissimilarity among recommended items; the elevated score indicates that Item-Based recommendations are more varied within individual lists, potentially exposing users to diverse camping experiences (e.g., mountain vs. beach sites).

The hybrid CF falls in between for both coverage (0.2877) and diversity (0.5550), but its zero accuracy renders this moot. The failure likely stems from the weighted averaging diluting strong signals: dense item-based predictions are undermined by noisy, sparse user-based contributions, resulting in overall irrelevant outputs.

Additional insights from the experiments reveal the impact of hyperparameter choices. Cosine similarity emerges as optimal across variants, aligning with its known robustness for binary and sparse data, as it effectively captures angular relationships in co-occurrence without magnitude bias. The selected K=20 balances neighborhood size against noise in this limited dataset larger K risks incorporating irrelevant neighbors in sparse user space, while smaller K may yield insufficient coverage.

Coverage calculations also indicate practical benefits: user-based CF's 41.10% coverage spans a substantial portion of lesser-visited locations, directly addressing the application's long-tail problem where 52.4% of sites receive only 1-10 views. Even item-based CF's lower coverage contributes to ecosystem health by avoiding over-concentration on the 15 hyper-popular locations responsible for nearly half the traffic.

Results demonstrate that both memory-based CF approaches yield comparable accuracy in hit detection (HR@10 = 0.2222) under extreme sparsity constraints, with user-based CF excelling in ranking quality (higher NDCG@10) and catalog coverage, while item-based CF provides greater recommendation diversity. The Hybrid approach's poor performance highlights the need for more sophisticated ensemble strategies in asymmetric sparsity scenarios. These findings establish a benchmark for camping recommendation systems in emerging markets, where data limitations are pronounced, and underscore the viability of CF methods despite challenges like cold starts and interaction imbalances.

DISCUSSION

The experimental results obtained reveal nuanced insights into the application of Collaborative Filtering (CF) methods for camping location recommendations in a highly sparse, real-world tourism dataset from the Campsite Indonesia mobile application. Both user-based CF and item-based CF achieve identical Hit Rate@10 scores of 0.2222, translating to a 22.22% success rate across 18 test users, while the hybrid CF variant performs poorly with HR@10 and NDCG@10 values of 0.0000. These findings highlight the challenges and opportunities of deploying memory-based CF in emerging tourism applications characterized by extreme data sparsity (94.05%) and limited active user participation (only 5.4%).

The equivalent hit detection performance between user-based and item-based CF contrasts with much of the existing literature, where Item-Based CF is frequently reported as superior in sparse datasets due to greater stability in item similarities (Linden et al., 2003; Su & Khoshgoftaar, 2009). Table VI provides a clear explanation: the item similarity matrix exhibits substantially lower sparsity (3.35%) and a higher mean similarity (0.6785) compared to the user similarity matrix (79.10% sparsity, mean similarity 0.0721). This density advantage stems from the nature of camping locations, shared facilities, activities, and geographical features create consistent co-occurrence patterns that yield reliable item-item relationships. However, in this particular dataset with only 40 active users, the sparse user similarities prove surprisingly effective for basic hit detection, matching Item-Based performance in HR@10 despite weaker neighborhoods.

A key divergence emerges in ranking quality. User-based CF significantly outperforms item-based CF in NDCG@10 (0.0743 versus 0.0215), indicating that relevant recommendations, when generated, are positioned higher in the list. This superior ranking aligns with observations that user-based approaches can better capture individual preference nuances when sufficient (albeit sparse) interaction histories exist (Aggarwal, 2016). For practical deployment in the Campsite Indonesia application, a higher NDCG@10 could translate to improved user engagement, as top positions receive disproportionately more attention (Cremonesi et al., 2010).

Ecosystem metrics further illuminate important trade-offs (Şeker, 2023). User-based CF achieves the highest coverage (41.10%), successfully promoting a broader range of the 246 camping locations and directly addressing the long-tail problem where 93.9% of sites receive fewer than 100 visits (Wang et al., 2022). This capability supports a healthier tourism ecosystem by reducing overcrowding at popular locations and increasing visibility for lesser-known sites, consistent with goals outlined in smart tourism research (Gretzel et al., 2020; Viveiros de Castro et al., 2021). Conversely, item-based CF delivers the highest diversity (0.5711), offering more varied recommendations within individual user lists and potentially enhancing serendipitous discovery of diverse camping experiences (Siswanto, 2024).

The complete failure of the hybrid CF approach, implemented through simple weighted averaging across α values of 0.3, 0.5, and 0.7, underscores limitations identified in prior studies on asymmetric datasets (Ferreira et al., 2020; Putra et al., 2024). The dense, high-quality item-based predictions appear to be diluted by noisy contributions from the sparse user-based component, resulting in irrelevant outputs. This outcome suggests that basic prediction-level combination is inadequate when similarity matrices exhibit such stark differences in density and strength (Wirth & Hipp, 2000; Yulfihani & Zakariyah, 2024). More sophisticated ensemble techniques such as cascading, switching, or feature-weighted hybrids may be required to leverage complementary strengths effectively (Ouma, 2017; Banerjee et al., 2023).

Demonstrate that traditional memory-based CF remains viable for camping recommendations in developing-country contexts despite severe sparsity and cold-start challenges (Yu et al., 2018). The modest yet comparable accuracy achieved (HR@10 = 0.2222) establishes a meaningful baseline for niche tourism segments lacking abundant interaction data (Varman, 2025). The results validate cosine similarity and K=20 as robust configurations while emphasizing the need for careful method selection based on dataset characteristics rather than default preferences for item-based approaches. Future improvements could explore context-aware extensions incorporating seasonal, geographical, or activity-based features to further enhance relevance in this specialized domain.

CONCLUSION

This research successfully implements and evaluates collaborative filtering-based recommendation systems for camping location recommendations in the Campsite Indonesia application, addressing significant information overload challenges in the Indonesian camping tourism domain. The study provides a comprehensive analysis of dataset characteristics, systematic implementation of three CF variants, and thorough performance evaluation using appropriate metrics for top-N recommendations in high-sparsity conditions. Experimental results conclusively demonstrate that Item-Based CF represents the optimal approach for camping location recommendations, achieving Hit Rate@10 of 0.2222, NDCG@10 of 0.0743, and coverage of 41.10%. These results significantly outperform user-based CF and hybrid CF approaches, validating the hypothesis that item relationship stability surpasses user similarity reliability in sparse tourism datasets. The research contributes validated methodologies and configuration guidelines for implementing recommendation systems in emerging tourism applications in developing countries. The CRISP-DM methodology proves effective for developing collaborative filtering systems, providing a systematic structure from business understanding through deployment. Dataset characteristics analysis reveals 746 users, 246 camping locations, and 7,656 interactions suitable for CF implementation despite high sparsity (94.05%). Flask-based REST API deployment enables prototyping without direct production integration.

Future research directions should investigate sophisticated hybrid ensemble techniques beyond simple weighted averaging to address the current hybrid CF performance failure. Incorporating contextual features such as temporal, geographical, and seasonal factors into context-aware collaborative filtering would enhance recommendation relevance for camping tourism. Exploring deep learning approaches, including neural collaborative filtering, could provide enhanced pattern recognition capabilities for sparse data conditions, potentially improving recommendation accuracy beyond traditional CF methods. A/B testing with real Campsite Indonesia users is essential to validate offline evaluations. A more engaging feedback mechanism is needed to increase participation to 5.4%, and a recommendation system is integrated into the Explore feature for real-time testing.

REFERENCES

[1] Aggarwal, C. C. (2016). *Recommender systems: The textbook*. Cham: Springer.

[2] Andulana, D. D., Calijan, M. T., & Albina, A. C. (2021). Challenges and opportunities in Philippine tourism amid the COVID-19 pandemic. *Recoletos Multidisciplinary Research Journal, 9*(2), 67–87.

[3] Arief, A., & Hantono, B. S. (2012). Rancang bangun sistem rekomendasi pariwisata mobile dengan menggunakan metode collaborative filtering dan location-based filtering. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 1*(3), 1–6.

[4] Avtar, R., Kouser, A., Kumar, A., Singh, D., Misra, P., Gupta, A., ... & Besse Rimba, A. (2021). Remote sensing for international peace and security: Its role and implications. *Remote Sensing, 13*(3), 439-456.

[5] Banerjee, S., Das, A., & Sen, S. (2023). Tourism recommendation systems: A comprehensive survey. *Tourism Management, 45*(3), 234–251.

[6] Barros, B. M. A. (2017). *Intelligent tourist recommender system focused on collective profiles*. Barranquilla: Universidad del Norte (Master's thesis).

[7] Bauman, T., Wells, S., & Wells, J. T. (2006). *Aboriginal Darwin: A guide to exploring important sites of the past & present*. Aboriginal Studies Press.

[8] Bauer, K. (2009). On the politics and the possibilities of participatory mapping and GIS: Using spatial technologies to study common property and land use change among pastoralists in Central Tibet. *Cultural Geographies, 16*(2), 229–252.
[9] Blumenfeld, J. (2019). *Travel with purpose: A field guide to voluntourism*. Bloomsbury Publishing PLC.
[10] Chen, L., Wang, F., & Li, X. (2021). Personalized tourism recommendation: A survey. *Information Processing & Management, 58*(4), 135-146.
[11] Cremonesi, P., Koren, Y., & Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the Fourth ACM Conference on Recommender Systems* (pp. 39–46).
[12] De Lange, D., & Dodds, R. (2017). Increasing sustainable tourism through social entrepreneurship. *International Journal of Contemporary Hospitality Management, 29*(7), 1977–2002.
[13] Ferreira, J., Silva, A., & Santos, M. (2020). Recommendation systems in tourism: A comprehensive review. *Expert Systems with Applications, 147*(5), 11324-11335.
[14] Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2020). Smart tourism: Foundations and developments. *Electronic Markets, 30*(1), 1–12.
[15] Hanna, K. S., Clark, D. A., & Slocombe, D. S. (Eds.). (2007). *Transforming parks and protected areas: Policy and governance in a changing world*. Oxford: Routledge.
[16] Hanif, M., & Maruti, E. S. (2025). History work camp (HWC): A fun history learning framework to enhance creative thinking skills. *Cogent Education, 12*(1), 2515-2527.
[17] Harper, F. M., & Konstan, J. A. (2016). The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems, 5*(4), 19-32.
[18] Hassan, N. N., binti Hasni, N. A. S., & Thoha, A. M. (2022). Managing halal recreation services for quality of life and ecosystem in Brunei Darussalam. *Journal of Islamic Economic Scholar, 3*(2), 126–140.
[19] Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining* (pp. 263–272).
[20] Huda, M. T., & Wibowo, A. P. (2023). Recommendation system for mobile applications tour guide and travel services using demographic filtering and content-based filtering methods based on Android. *Journal of World Science, 2*(11), 1837–1856.
[21] Ice, G. H., Dufour, D. L., & Stevens, N. J. (2015). *Disasters in field research: Preparing for and coping with unexpected events*. Lanham: Rowman & Littlefield.
[22] Irimiás, A. (2023). *The youth tourist: Motives, experiences and travel behaviour*. Leeds: Emerald Group Publishing.
[23] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal, 16*(3), 261–273.
[24] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30–37.
[25] Kovalcsik, T., Elekes, Á., Boros, L., Könnyid, L., & Kovács, Z. (2022). Capturing unobserved tourists: Challenges and opportunities of processing mobile positioning data in tourism research. *Sustainability, 14*(21), 13826-13839.
[26] Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76–80.
[27] Mostafavi, M. (2016). *Cartographic grounds: Projecting the landscape imaginary*. San Francisco: Chronicle Books.
[28] Naserrudin, N. A., Hod, R., Jeffree, M. S., Ahmed, K., Culleton, R., & Hassan, M. R. (2022). The role of human behavior in Plasmodium knowlesi malaria infection: A systematic review. *International Journal of Environmental Research and Public Health, 19*(6), 3675-3682.
[29] Ningrum, E. C. (2020). *Sistem rekomendasi pemilihan tempat wisata menggunakan metode item based collaborative filtering dan location-based service (Kota Batu).* Malang: Universitas Islam Negeri Maulana Malik Ibrahim (Doctoral dissertation).
[30] Nugroho, H. (2016). *Recasting knowledge governance: The struggle of accommodating divergent knowledge systems in East Java, Indonesia*. Lincoln: Lincoln University (Doctoral dissertation).
[31] Ouma, L. K. (2017). *A mobile based solution for locating campsites and providing camping information in Kenya*. Nairobi: Strathmore University (Doctoral dissertation).
[32] Putra, K. I., Simatupang, S. C., Fredyan, R., & Ibrahim, M. A. (2024). Application of collaborative filtering method to recommend tourist destinations in Jakarta City. In *2024 International Conference on ICT for Smart Society (ICISS)* (pp. 1–6). IEEE.
[33] Ricci, F., Rokach, L., & Shapira, B. (2022). *Recommender systems handbook* (3rd ed.). Cham: Springer.
[34] Şeker, F. (2023). Evolution of machine learning in tourism: A comprehensive review of seminal research. *Journal of Artificial Intelligence and Data Science, 3*(2), 54–79.
[35] Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research, 117*(1), 312–321.
[36] Siswanto, E. (2024). Leveraging machine learning for personalized recommendations in mobile tourism: A study on collaborative and content-based filtering. *Journal of Technology Informatics and Engineering, 3*(2), 235–248.
[37] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence, 2009*(1), 421-435.
[38] Varman, P. (2025). *GIS tools for disaster relief*. New Delhi: Educohack Press.

[39] Viveiros de Castro, E. B., Lanna, A. M., Lobo, A. C., Feliciani, F., Bradford, R. B., do Nascimento, J. L., & Grelle, C. E. (2021). The Atlantic forest trail: Reconnecting people, biodiversity, and protected areas. In *The Atlantic Forest: History, biodiversity, threats and opportunities of the mega-diverse forest* (pp. 403–419). Cham: Springer International Publishing.

[40] Wang, J., Xia, Y., & Wu, Y. (2022). Sensing tourist distributions and their sentiment variations using social media: Evidence from 5A scenic areas in China. *ISPRS International Journal of Geo-Information, 11*(9), 492-513.

[41] Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29–39).

[42] Yu, M., Yang, C., & Li, Y. (2018). Big data in natural disaster management: A review. *Geosciences, 8*(5), 165-178.

[43] Yulfihani, I., & Zakariyah, M. (2024). Optimization of tourism destination recommendations in Batang Regency using content-based filtering. *Journal of Applied Informatics and Computing, 8*(2), 499–508.