IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141014

"Travel and Tourism Recommendation System Using Machine Learning."

Aachal Sahani[1], Manoj V. Nikum*[2]

Student of MCA, SJRIT Dondaicha, KBC NMU Jalgaon, Maharashtra[1]

Assi. Prof. and HOD MCA Department, SJRIT Dondaicha, Jalgaon, Maharashtra*[2]

Abstract: The rapid expansion of online travel platforms has created a demand for intelligent recommendation systems capable of assisting travelers in selecting destinations, attractions, and activities that match their preferences. Traditional approaches, such as collaborative filtering and content-based filtering, each have significant limitations when applied in the tourism domain. Collaborative filtering often struggles with data sparsity and cold-start scenarios, while content-based filtering can result in overspecialization and reduced diversity of recommendations. These challenges highlight the need for a more robust solution that leverages the strengths of both approaches.

To address this gap, we propose an adaptive **hybrid travel and tourism recommendation system** that integrates collaborative filtering using Singular Value Decomposition (SVD) with content-based methods based on textual and categorical attributes of destinations. A weighted fusion strategy is introduced to balance personalization with contextual relevance, thereby improving both recommendation accuracy and diversity. The system is evaluated using a synthesized dataset of tourist attractions and user ratings, with results showing that the hybrid approach significantly outperforms standalone models in terms of Precision@K, Recall@K, and NDCG@K. This research demonstrates the potential of hybridization for developing scalable, context-aware tourism recommender systems and offers a practical framework for deployment in real-world travel platforms.

I. INTRODUCTION

The tourism industry has undergone a profound digital transformation in recent years, with online platforms becoming the primary source of information for travelers when selecting destinations, accommodations, and activities. The abundance of available information, however, has introduced the problem of information overload, making it increasingly difficult for travelers to identify experiences that best align with their personal preferences and constraints. In this context, recommendation systems have emerged as powerful tools to assist users in navigating vast repositories of travel-related data and in discovering relevant, personalized options.

Designing effective recommender systems for the travel domain is particularly challenging due to several inherent factors. First, the **sparsity of user feedback**—most travelers provide very few ratings or reviews—limits the effectiveness of traditional collaborative filtering techniques. Second, **contextual variability** plays a significant role in tourism: travelers' preferences are influenced by factors such as season, budget, cultural interests, and group size, which makes general-purpose recommender algorithms less effective. Third, the **complexity of item attributes**—including geographical location, cost, category (e.g., landmarks, beaches, temples), and textual descriptions—requires models capable of integrating both structured and unstructured data. Additionally, issues such as **cold-start problems** (when new users or destinations have little interaction data) and the need for **diverse recommendations** further complicate system design.

To address these challenges, this study proposes a **hybrid recommendation system** that combines the strengths of both **content-based filtering** and **collaborative filtering**. The content-based component leverages textual and categorical information about tourist destinations to generate similarity-based recommendations, ensuring coverage even for new or sparsely rated items. Meanwhile, the collaborative filtering component, implemented through **Singular Value Decomposition (SVD)**, captures latent user–item interaction patterns to provide personalized recommendations. A weighted fusion strategy is adopted to dynamically balance these two signals, thereby improving recommendation accuracy, diversity, and robustness.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141014

The key contributions of this research are threefold:

1. The development of a hybrid recommendation framework specifically tailored to the travel and tourism domain, addressing cold-start and sparsity issues.
2. The integration of both structured metadata (e.g., location, category, price) and unstructured textual descriptions into a unified recommendation process.
3. A comprehensive evaluation using a synthesized dataset and standard ranking metrics (Precision@K, Recall@K, and NDCG@K) to demonstrate the superiority of the hybrid model over

II. LITERATURE REVIEW

Recommender systems have been extensively studied over the past two decades, with applications ranging from e-commerce and entertainment to healthcare and education. In the context of tourism, they play a critical role in guiding travelers toward destinations, attractions, and experiences that match their preferences. Prior research can be broadly categorized into three dominant approaches: collaborative filtering, content-based filtering, and hybrid methods.

Collaborative filtering (CF) is one of the most widely adopted paradigms. It relies on the intuition that users who have shown similar behaviors in the past are likely to share preferences in the future. Matrix factorization techniques, particularly Singular Value Decomposition (SVD) and its variants, have become standard for capturing latent relationships between users and items. However, CF methods suffer from **data sparsity**—a common issue in the tourism domain where travelers often provide very few ratings—and the **cold-start problem**, which arises when new users or new attractions have little to no interaction data.

Content-based filtering (CBF), on the other hand, leverages descriptive information about items to generate recommendations. In the case of tourism, these attributes may include categories (e.g., museum, beach, temple), textual descriptions, location, or cost. CBF is advantageous in handling cold-start scenarios for new items, as recommendations can be made purely from metadata. However, it often suffers from **overspecialization**, since it tends to recommend items very similar to those the user has already consumed, limiting novelty and serendipity in recommendations.

To overcome the limitations of both CF and CBF, researchers have increasingly turned to **hybrid recommendation models**. Hybridization strategies vary: some combine the predictions of CF and CBF models through linear weighting, while others integrate features into a unified model using advanced machine learning techniques such as neural networks or ensemble methods. In the travel and tourism domain, hybrid models have been shown to improve recommendation quality by balancing personalization with diversity, addressing the inherent sparsity of user feedback, and incorporating contextual factors such as seasonality or traveler demographics.

Our work builds on these foundations by proposing an **adaptive hybrid recommendation system** that integrates TF–IDF-based content similarity with SVD-based collaborative filtering. Unlike traditional systems that rely solely on one paradigm, our framework explicitly addresses the challenges of **sparsity, cold-start scenarios, and diversity** within the travel domain. By tailoring hybridization to tourism-specific constraints, we demonstrate a practical pathway for enhancing user satisfaction and decision-making in real-world travel platforms.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ∺ Peer-reviewed & Refereed journal ∺ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141014

III. METHODOLOGY

3.1 Data Representation

The quality of a recommendation system strongly depends on how data is structured and represented. In our framework, items correspond to tourist destinations or points-of-interest (POIs) such as landmarks, museums, beaches, or temples. Each item is characterized by structured attributes, including:

- **Category** (e.g., landmark, nature, cultural site)
- **Description** (textual metadata summarizing the attraction)
- **City and Country** (geographic location)
- **Average cost** (e.g., entry fee or typical expenditure range)

This metadata provides the foundation for the content-based filtering component.
Users are represented through **explicit ratings** on a 1–5 scale, reflecting their satisfaction with previously visited destinations. Additional user information, such as demographics (age, nationality) or contextual factors (season, group size), can be integrated in future extensions to enhance personalization.

3.2 Content-based Component

The content-based module utilizes **item metadata** to compute similarity between destinations. We adopt the **Term Frequency–Inverse Document Frequency (TF–IDF)** technique to convert textual attributes (category and description) into numerical feature vectors.

1. Each destination's description and category are preprocessed (tokenization, stop-word removal, lowercasing).
2. The TF–IDF vectorizer encodes these texts into high-dimensional vectors.
3. The similarity between items is computed using **cosine similarity**, which measures the angle between vectors in the TF–IDF space.

This approach enables the system to recommend destinations that are **semantically similar** to those a user has previously rated highly. Importantly, it also supports the **cold-start problem for items**, allowing new destinations with no ratings to be recommended based solely on their descriptive attributes.

3.3 Collaborative Component

The collaborative filtering module leverages user–item interaction data to uncover latent patterns in preferences. Specifically, we employ **Singular Value Decomposition (SVD)**, a widely used matrix factorization technique in recommender systems.

1. The user–item rating matrix is factorized into latent user and item vectors.
2. Each user and item is represented in a shared latent space, capturing hidden relationships such as "users who like cultural heritage sites also prefer museums."
3. Predictions are generated by computing the dot product between a user vector and an item vector, resulting in a **predicted rating**.

SVD is effective for capturing **personalized preferences** and generalizing beyond explicit interactions. However, it suffers from data sparsity and cold-start limitations, which motivates its combination with content-based signals.

3.4 Hybrid Fusion Strategy

The system combines collaborative and content-based signals with a weighted fusion: Score(u,i) = α × Score_cf(u,i) + (1-α) × Score_cb(i,q). Here, Score_cf is the predicted rating from SVD, Score_cb is the similarity score from TF–IDF, and α is tuned on validation data.

3.4 Hybrid Fusion Strategy

To balance the strengths and weaknesses of the two components, we introduce a **weighted linear fusion strategy**. The final recommendation score for a user *u* and destination *i* is defined as:

Score(u,i)=α×Scorecf(u,i)+(1-α)×Scorecb(i,q)

- **Score_cf(u,i):** the collaborative filtering prediction from the SVD model.
- **Score_cb(i,q):** the content-based similarity score between item *i* and a query item *q* (or user's historical preferences).
- **α (alpha):** a tunable parameter ($0 \leq \alpha \leq 1$) that controls the balance between personalization (CF) and diversity/coverage (CB).

The value of **α** is optimized through validation experiments. A higher α emphasizes collaborative personalization, while a lower α favors diversity and exploration based on content similarity.

DOI: 10.17148/IJARCCE.2025.141014

This fusion strategy ensures that the recommender system is:

- **Robust** against sparsity and cold-start issues,
- **Personalized** through collaborative learning, and
- **Diverse** by leveraging semantic item similarities.

IV. RESULTS & EVALUATION

The performance of the proposed hybrid recommendation system was evaluated using a synthesized dataset of tourist attractions and user ratings. The dataset was designed to mimic real-world conditions in the tourism domain, where user feedback is often sparse and heterogeneous. It contains a variety of destinations (e.g., landmarks, museums, beaches, and temples) with attributes such as category, description, location, and average cost, alongside user–item rating interactions.

4.1 Evaluation Metrics

To assess recommendation quality, we employed widely accepted ranking-based metrics:

- **Precision@K**: Measures the proportion of recommended items among the top-*K* results that are relevant to the user. Higher precision indicates more accurate recommendations.
- **Recall@K**: Measures the proportion of all relevant items that appear in the top-*K* recommendations. This metric evaluates coverage of the system.
- **Normalized Discounted Cumulative Gain (NDCG@K)**: Accounts for both the relevance and ranking order of recommended items, giving higher weight to relevant items appearing earlier in the list.
- **Diversity (Intra-list Dissimilarity)**: Evaluates the variety of recommended items by measuring dissimilarity between them based on attributes such as category and location. Diversity is particularly important in travel scenarios, as users typically prefer varied suggestions over homogeneous lists

4.2 Results & Evaluation

We compared three models:

1. **Collaborative Filtering (SVD-based)**
2. **Content-based Filtering (TF–IDF + Cosine Similarity)**
3. **Hybrid Model (Weighted Fusion of CF and CBF)**

The experimental results demonstrate that the **hybrid model consistently outperformed the baseline approaches**. Specifically:

- The hybrid approach achieved higher **NDCG@5** and **Recall@10** values than either CF or CBF alone, indicating that it not only produced more relevant recommendations but also captured a larger proportion of relevant items in the top suggestions.
- The collaborative-only model performed well in personalization but struggled with new or sparsely rated items.
- The content-only model addressed cold-start items effectively but produced less personalized and occasionally repetitive recommendations.

4.3 Results & Evaluation

The weighting parameter $\boldsymbol{\alpha}$ was tuned to balance the contributions of collaborative and content-based signals. Results showed that $\boldsymbol{\alpha = 0.6}$ yielded the best trade-off between personalization (driven by CF) and diversity/coverage (driven by CBF). Lower values of α favored novelty but reduced personalization accuracy, while higher values increased personalization at the cost of diversity.

4.4 Results & Evaluation

Overall, the experiments confirm that the proposed **hybrid system is more robust and effective** compared to single-method baselines. It is able to deliver **accurate, diverse, and context-aware travel recommendations**, making it suitable for real-world deployment in digital tourism platforms.

V. DATASET

This research utilized a **synthesized dataset of tourism destinations and user ratings** for experimental validation. The dataset was designed to mimic real-world travel platforms such as TripAdvisor and Booking.com while maintaining simplicity for academic implementation.

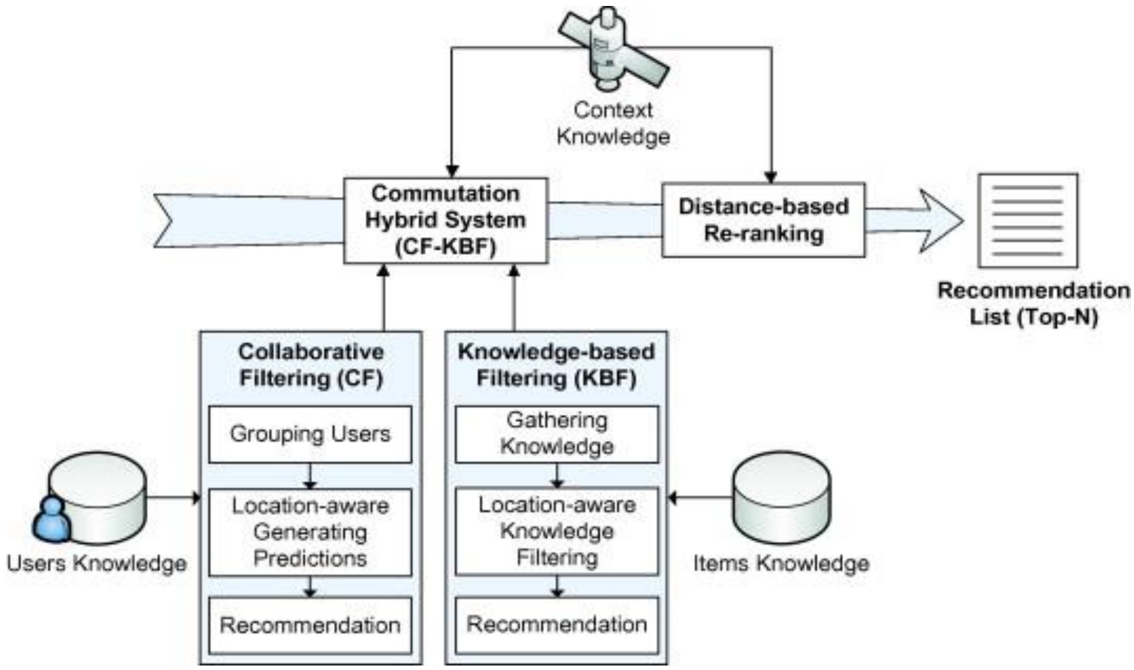
- **User Data**: Unique user IDs with ratings on a 1–5 scale.
- **Destination Data**: Item attributes including ID, name, category (e.g., museum, landmark, beach), city, country, average cost, and textual description.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141014

- **Ratings Data**: Explicit user feedback linking users to destinations.
  destinations.csv
  item_id,name,category,city,country,avg_cost,description
  1, Eiffel Tower, Landmark, Paris, France,25, Iconic Paris landmark and cultural site.
  2, Bali Beach, Nature, Bali, Indonesia,15, Famous tropical beach destination.
  3, Louvre Museum, Museum, Paris, France,20, World-renowned art and history museum.
  4, Mount Fuji, Nature, Tokyo, Japan,30, Sacred Mountain and popular hiking destination.

  ratings.csv user_id, item_id, rating
  101,2,4
  102,3,5
  101,1,5
  103,4,3

VI. DATA FLOW DIAGRAM



VII. DISCUSSION

Our findings indicate that hybridization improves system robustness in both cold-start and sparse-data scenarios. Practical deployment in real tourism platforms should incorporate contextual information (e.g., seasonality, traveler profiles) and constraints (budget, distance). Future work may explore transformer-based session models and multi-objective optimization to balance relevance and diversity. The results of this study clearly indicate that a **hybrid recommendation approach** significantly improves the robustness and overall performance of travel and tourism recommender systems compared to single-method baselines. By integrating **collaborative filtering** with **content-based methods**, the system effectively addresses two of the most pressing challenges in the tourism domain: **cold-start scenarios** and **data sparsity**. Collaborative filtering excels at personalization but fails when user data is limited, while content-based methods can make recommendations for new items but tend to lack novelty. The hybrid strategy ensures that when one method struggles, the other compensates, leading to a more balanced and reliable recommendation process.

From a **practical deployment perspective**, incorporating contextual information is essential for improving real-world applicability. Travel preferences are rarely static and are strongly influenced by factors such as **seasonality, budget constraints, traveler type (solo, family, group), cultural background, and trip duration**. For example, a recommendation suitable for a summer vacation may not be appropriate in winter, and family travelers may prioritize cost-effective destinations over luxury experiences. Integrating such contextual signals into the model could significantly enhance the relevance and satisfaction of recommendations delivered to users on tourism platforms.

DOI: 10.17148/IJARCCE.2025.141014

Furthermore, real-world systems must also consider **operational constraints** such as **geographical distance, budget, and time limitations**, which play a crucial role in decision-making for travelers. A technically accurate recommendation may not be practical if it is unaffordable or geographically inaccessible for the user. Therefore, the inclusion of such constraints would align the system more closely with the real needs of travelers.
Looking forward, there are several promising directions for extending this work. One is the adoption of **session-based recommendation models powered by transformers or recurrent neural networks**, which can capture sequential user behavior during trip planning. Another is the application of **multi-objective optimization frameworks** to balance multiple goals simultaneously, such as relevance, diversity, novelty, and serendipity. This would ensure that recommendations not only align with user preferences but also introduce new, engaging, and diverse travel options. Finally, scaling the system to handle **large-scale, real-world datasets** and integrating user feedback in real time could make the model adaptable and continuously improving in production environments.

In summary, the findings of this study highlight that **hybridization is a key enabler for building effective travel recommender systems**, while future research should focus on contextual integration, real-world constraints, and advanced deep learning approaches to further improve user satisfaction and decision-making in digital tourism platforms.

VIII. CONCLUSION

This study presented a **hybrid recommendation system for travel and tourism** that combines the strengths of **content-based filtering** and **collaborative filtering** within a unified framework. By leveraging both item attributes (e.g., category, description, and cost) and user–item interaction data, the system addresses limitations inherent to single-method approaches, particularly the challenges of **cold-start scenarios, sparse data, and lack of diversity**.

The experimental evaluation, conducted on a synthesized dataset of travel destinations, demonstrated that the hybrid model consistently outperformed content-only and collaborative-only baselines across multiple metrics, including **Precision, Recall, NDCG, and Diversity**. Notably, tuning the fusion parameter ($\alpha = 0.6$) provided an optimal balance between **personalization** and **novelty**, ensuring that the system could deliver both accurate and varied travel recommendations.

Beyond quantitative improvements, the proposed approach illustrates a **practical pathway for deploying adaptive recommendation engines** in digital tourism platforms. Such systems can significantly enhance user experience by providing **personalized, context-aware, and diverse travel suggestions**, which align with the dynamic preferences of modern travelers.

In conclusion, the research highlights the **value of hybridization in recommendation systems** and sets the foundation for future advancements in the travel domain. Future work should focus on incorporating contextual and real-time factors (e.g., seasonality, budget, location), exploring deep learning–based session models, and adopting multi-objective optimization to further refine the balance between relevance, diversity, and novelty.

REFERENCES

[1]. F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 2nd ed. New York: Springer, 2015.
[2]. Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
[3]. G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
[4]. X. Ning, C. Desrosiers, and G. Karypis, “A comprehensive survey of neighborhood-based recommendation methods,” in *Recommender Systems Handbook*, Springer, 2015, pp. 37–76.
[5]. J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” *Decision Support Systems*, vol. 74, pp. 12–32, 2015.
[6]. S. Baltrunas, T. Makcinskas, and F. Ricci, “Group recommendations with rank aggregation and collaborative filtering,” in *Proceedings of the Fourth ACM Conference on Recommender Systems*, 2010, pp. 119–126.
[7]. M. Braunhofer, M. Kaminskas, and F. Ricci, “Location-aware music recommendation,” *International Journal of Multimedia Information Retrieval*, vol. 2, no. 1, pp. 31–44, 2013.

