Enhanced Travel Recommendation System Using Hybrid Filtering

1st Dr.Gnanasekar J.M
Dept.of Artificial Intelligence and Data
Rajalakshmi Engineering College
Chennai, India

2nd Devika c
Dept.of Artificial Intelligence and Data science science
Rajalakshmi Engineering College
Chennai, India
221801009@rajalakshmi.edu.in

3rd Keerthika P
Dept.of Artificial Intelligence and Data science
Rajalakshmi Engineering College
Chennai, India
221801027@rajalakshmi.edu.in

Abstract— **In today's digital age, travelers increasingly depend on internet platforms to plan their vacations, generating vast amounts of data that can be leveraged to enhance the recommendation process. However, existing Travel Recommendation Systems (TRS) often suffer from limitations such as data sparsity, resulting in suboptimal recommendations when relying solely on collaborative or content-based filtering methods. This paper presents a novel travel booking website that addresses these limitations by implementing a hybrid filtering approach, combining collaborative filtering and content – based filtering .The collaborative filtering component utilizes the Singular Value Decomposition (SVD) algorithm to analyze user rating patterns and predict travel destination preferences, even for users with limited data. Meanwhile, the content-based filtering component leverages the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm to analyze textual reviews and recommend destinations based on their similarity to previously visited places.**

The system further enhances the recommendation process by integrating both approaches through an adaptive weighted scoring mechanism, ensuring that recommendations are not only highly relevant but also up-to-date with user preferences and evolving content. This hybrid system effectively mitigates the data sparsity problem by drawing on both explicit user feedback (ratings) and implicit feedback (reviews), leading to more accurate and diverse travel suggestions. By continuously refining its recommendations based on new data and user behavior, the proposed system significantly improves the user experience, providing a dynamic and enriched travel planning tool. The combination of collaborative and content-based filtering in a hybrid model offers a robust solution to the challenges faced by traditional TRS, making it an ideal choice for personalized travel suggestions in the modern digital era.

Keywords— Travel Recommendation System, Hybrid Filtering, Collaborative Filtering, Singular Value Decomposition (SVD), Content-Based Filtering, Term Frequency-Inverse Document Frequency (TF-IDF), Personalized Travel Recommendations, User Experience.

I. INTRODUCTION

The travel industry has seen a paradigm shift with the advent of digital platforms and the growing demand for personalized experiences. Modern travelers no longer seek generic recommendations but instead expect tailored suggestions that cater to their unique preferences and requirements, ranging from ideal destinations to specific activities and accommodations. Recommendation systems have become an indispensable tool for achieving this personalization, enhancing user satisfaction and engagement on platforms like Airbnb, TripAdvisor, and Expedia.

Despite significant advancements, traditional recommendation systems face persistent challenges that limit their effectiveness. Key issues include **data sparsity** and the **cold-start problem**. Data sparsity arises when there is insufficient user-item interaction data to generate reliable recommendations, while the cold-start problem refers to the difficulty of making accurate recommendations for new users or less popular destinations due to a lack of prior data. These challenges are particularly pronounced in the travel domain, where user preferences are influenced by diverse factors such as location, climate, budget, and activity type.

Traditional approaches, such as **collaborative filtering** and **content-based filtering**, address these issues to some extent but exhibit significant limitations. Collaborative filtering relies on the preferences of similar users, making it effective for identifying trends within user groups. However, its reliance on historical interaction data makes it inadequate for new users or items. Content-based filtering, on the other hand, uses attributes of destinations or items to generate recommendations. While this approach excels at ensuring relevance, it often fails to promote discovery of new or unconventional experiences, limiting its utility in encouraging exploratory behavior.

To overcome these challenges, this project proposes a **hybrid recommendation system** that integrates collaborative filtering and content-based filtering. By leveraging the strengths of both approaches, the hybrid system provides a balanced solution that enhances personalization while addressing the cold-start problem and data sparsity. Collaborative filtering enables the system to recommend novel destinations based on user similarities, while content-based filtering ensures that recommendations align with individual preferences using specific destination attributes such as location, activity type, and budget.

The proposed system is designed to adapt and improve over time by learning from user interactions, thereby refining its recommendations dynamically. It leverages content-based filtering to provide meaningful suggestions for new users and unexplored destinations, ensuring that even with minimal interaction data, the system delivers value. As user data accumulates, collaborative filtering further enhances the personalization and diversity of recommendations.

This hybrid approach has the potential to transform the travel planning experience by providing accurate, diverse, and engaging recommendations. It aims to improve user satisfaction, increase engagement, and drive higher conversion rates for travel platforms. Moreover, the insights gained from this project contribute to the broader field of recommendation systems, offering a scalable and adaptable framework that can be applied to other domains such as e-commerce, hospitality, and entertainment. By addressing critical challenges in recommendation systems, this project seeks to redefine how users discover and plan their travels, paving the way for more intuitive and user-centric platforms.

II. RELATED WORKS

The advancements in travel recommendation systems have significantly improved user experiences, but challenges like data sparsity and the cold-start problem persist. Numerous studies have explored hybrid recommendation systems that combine collaborative filtering (CF) and content-based filtering (CBF) techniques to mitigate these issues. This section elaborates on relevant literature addressing the proposed hybrid travel recommendation system.

1. Hybrid Recommendation Techniques

Several studies emphasize the efficacy of hybrid systems in overcoming the limitations of traditional CF and CBF. According to [1], hybrid models leverage CF's ability to discover patterns in user interactions while using CBF to integrate item-specific attributes like destination features and activities. This dual approach addresses sparsity by ensuring meaningful recommendations, even for new users or unexplored items.

2. Addressing Cold-Start and Sparsity

The cold-start problem, arising from insufficient interaction data, is a major bottleneck for recommendation systems. Research in [2] demonstrates hybrid systems using auxiliary data like user demographics or contextual information to initialize recommendations. Content-driven methods are often employed initially, transitioning to collaborative methods as user interaction data accumulates.

3. Role of Big Data and Machine Learning

Big data analytics and machine learning (ML) play pivotal roles in enhancing hybrid recommendation systems. Studies such as [3] detail the application of deep learning algorithms to extract meaningful insights from user-generated content, such as reviews and ratings. By integrating these insights into hybrid frameworks, systems achieve greater personalization and adaptability.

4. Multi-Objective Optimization

In [4], a multi-objective hybrid approach was explored, balancing accuracy with diversity in recommendations. Such frameworks optimize user satisfaction while encouraging the discovery of novel destinations, addressing the often-overlooked need for varied recommendations.

5. User Experience and Personalization

Improving user engagement and personalization remains a primary focus. According to [5], hybrid models outperform standalone CF and CBF methods in terms of precision, recall, and diversity metrics. Tailored travel recommendations are particularly effective in increasing booking conversions and fostering user trust.

6. Applications in Tourism

Hybrid recommendation systems have also been applied to destination suggestions, itinerary planning, and activity recommendations. Research in [6] highlights the use of hybrid systems in identifying emerging travel trends, enabling platforms to remain competitive and relevant.

III. PROPOSED SYSTEM

The proposed travel recommendation system integrates **collaborative filtering (CF)** and **content-based filtering (CBF)** into a hybrid model to provide personalized and highly accurate travel recommendations. By leveraging both user interaction data and item attributes, it addresses common challenges like the cold start problem, data sparsity, and lack of personalization. The system continuously updates user profiles based on real-time interactions, including explicit feedback (ratings, bookings) and implicit feedback (browsing behavior). This allows the system to offer increasingly relevant recommendations over time. To enhance diversity, the system occasionally introduces travel options outside a user's typical preferences, promoting exploration. The hybrid filtering approach ensures that even with minimal user data, the content-based method provides meaningful suggestions, while collaborative filtering improves recommendations as more user data becomes available. Additionally, the system's scalability is supported by efficient algorithms like matrix factorization and TF-IDF, enabling it to handle large datasets of users and destinations. Ultimately, the proposed system delivers a personalized, diverse, and scalable travel recommendation experience, optimizing user satisfaction and engagement.

System Architecture

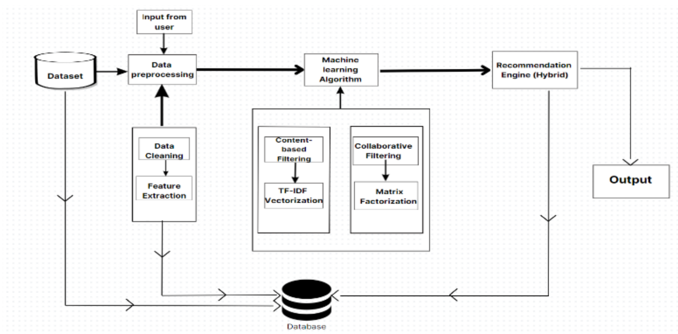

Fig. 1. System Architecture

Collaborative Filtering (CF) is a popular technique used in recommendation systems, including for travel platforms, to provide personalized recommendations based on user behavior and interactions. In collaborative filtering, the system analyzes the past behaviors of users, such as their travel destination preferences, bookings, ratings, or reviews, and identifies patterns among them. The basic idea is to recommend destinations to a user based on the preferences of similar users.

There are two main types of collaborative filtering: **user-based** and **item-based**. In **user-based collaborative filtering**, the system compares a user's preferences with those of other users who have similar interests and recommends destinations that those similar users liked. In **item-based collaborative filtering**, the system compares the destinations themselves—finding similarities between the destinations a user has interacted with and other destinations that are often liked by similar users. Collaborative filtering is particularly useful in travel recommendation systems as it allows users to discover new destinations they may not have considered, by leveraging the collective wisdom of other users with similar interests. However, one limitation of CF is the **cold start problem**, where new users or items with no prior interaction data may not receive accurate recommendations. This challenge can be mitigated by combining collaborative filtering with content-based filtering in a hybrid system.

A.Data Preprocessing

Data preprocessing is an essential step to ensure the accuracy and efficiency of the recommendation system. In this project, data preprocessing involves the following stages to prepare both user interaction and destination attribute data for analysis:

Data Collection: Data sources include user demographic information (e.g., age, gender, location), past interactions (e.g., bookings, ratings), and destination characteristics (e.g., type, activities, amenities).

Data Cleaning: This step handles missing values, outliers, and duplicate records to improve data quality. Imputation techniques are applied to fill in missing values, while irrelevant or redundant records are removed to ensure the consistency of input data across the recommendation system.

Feature Extraction and Scaling: Relevant features are extracted to create a streamlined dataset for the recommendation algorithms. Scaling ensures that numerical features are on a comparable scale, preventing dominance by attributes with larger ranges.

B.Collaborative Filtering using Singular Value Decomposition (SVD) Algorithm

Collaborative Filtering (CF) leverages user interaction data to make recommendations based on user or item similarities. ***Singular Value Decomposition (SVD)*** is a matrix factorization method applied within CF to capture latent user preferences, effectively overcoming challenges posed by data sparsity and cold start issues. An interaction matrix ***R*** is created, where rows represent users and columns represent items (*destinations*). Each cell in ***R*** indicates the interaction level, such as ratings The matrix ***R*** is factorized into three matrices***:*** ***U*** (*user feature matrix*), ***Σ*** (*diagonal matrix of singular values*), and ***V^T*** (*item feature matrix*), such that

$$\mathbf{R \approx U \Sigma V^T}$$

The matrices are truncated to retain only the top kkk singular values, resulting in U , Σ and V_k^T. This step reduces noise and focuses on the most informative latent factors, improving recommendation accuracy by concentrating on the primary interaction patterns.

The truncated matrices are recombined to create an approximate matrix

$$\mathbf{\hat{R} = U_k \Sigma_k V_k^T}$$

which includes estimated values for previously unobserved interactions.

For each user, the system recommends items (destinations) with the highest predicted interaction values in ***R^*** that the user has not yet interacted with.This approach is particularly useful for sparse datasets, as SVD effectively uncovers hidden relationships between users and items, providing meaningful recommendations even when data is limited.

C.Content based Filtering using TF-IDF

In recommendation system, **content-based filtering** is employed using **TF-IDF (Term Frequency-Inverse Document Frequency)** and **cosine similarity** to generate personalized recommendations based on the attributes of travel destinations. TF-IDF is used to convert textual data, such as destination descriptions and user reviews, into a numerical format that highlights important keywords, enabling the system to assess the relevance of each destination based on user interests. By computing the TF-IDF scores of different terms associated with a travel destination, the system can create a vector representation for each destination. To determine similarity between destinations and user preferences, **cosine similarity** is then applied to compare these vectors. Cosine similarity calculates the cosine of the angle between two vectors, measuring their similarity on a scale from 0 (no similarity) to 1 (identical). This method allows the system to recommend destinations that share similar characteristics to those a user has previously shown interest in, ensuring that recommendations are relevant to the user's preferences. Through the combination of TF-IDF and cosine similarity, the content-based filtering approach effectively suggests travel destinations that align with user profiles, even in the absence of collaborative data.

Term Frequency (TF)

Term Frequency measures how often a specific term appears in a document. It is calculated as:

$$\text{TF(t,d)} = \frac{\text{Number of times term t appears in document d}}{\text{Total number of terms in document d}}$$

Inverse Document Frequency (IDF)

While term frequency tells us how often a word appears in a specific document, it does not account for the importance of that word across the entire dataset. **Inverse Document Frequency** aims to solve this by measuring how unique or rare a term is across all documents.

IDF is calculated as:

$$\text{IDF(t)} = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing the term t}}\right)$$

TF-IDF Calculation:

TF-IDF(t,d)=TF(t,d)×IDF(t)

The **cosine similarity** formula is given by:

$$\text{Cosine Similarity}=\frac{A\cdot B}{\|A\|\|B\|}$$

Hybrid Filtering combines the strengths of multiple recommendation techniques, such as **Collaborative Filtering (CF)** and **Content-Based Filtering (CBF)**, to enhance the overall recommendation process. In the context of travel recommendations, the hybrid model integrates user behavior data (e.g., past bookings, ratings, and preferences) with item attributes (e.g., location type, cost, activities) to provide more accurate and personalized suggestions. This combination helps overcome the limitations of each individual method. For instance, while **Collaborative Filtering** excels in capturing user similarity and community-based patterns, it struggles with the **cold start problem** when dealing with new users or items. **Content-Based Filtering**, on the other hand, can recommend based on item attributes but may lack diversity in recommendations. The hybrid system alleviates these issues by leveraging **CBF** to provide recommendations when **CF** data is sparse or unavailable, and vice versa. As user interaction data grows, the system can increasingly rely on **Collaborative Filtering** to fine-tune recommendations, ensuring the system delivers high-quality, personalized suggestions regardless of the data available. By blending both methods, the hybrid filtering approach offers a balanced, scalable, and efficient recommendation engine for travel platforms, ensuring better coverage of user preferences and an enriched user experience.

Final Score=(w1×ScoreContent-Based)+(w2 ×ScoreCollaborative)

A **Recommendation Engine** in a travel system works by collecting user and item data to provide personalized suggestions. It begins by gathering information about user preferences, past interactions, and demographic details, alongside attributes of travel destinations, such as location, cost, and available activities. The system uses techniques like **Collaborative Filtering (CF)**, which identifies patterns based on user behavior and recommends destinations liked by similar users, and **Content-Based Filtering (CBF)**, which recommends destinations with attributes similar to those a user has previously engaged with. **Hybrid Filtering** combines both approaches to improve accuracy by leveraging the strengths of each method. Once the data is processed, the engine predicts which destinations the user is likely to prefer and ranks them based on factors like relevance and novelty. Recommendations are then delivered to the user, and the system continuously refines these suggestions by incorporating new user interactions, ensuring more accurate and personalized results over time. This process allows the recommendation engine to adapt and improve, offering a tailored travel experience to each user.

.Collaborative Filtering using Singular Value Decomposition (SVD) Algorithm

Step 1: Create the User-Item Rating Matrix

Construct a matrix RRR, where rows represent users and columns represent items (e.g., travel destinations). Matrix cells contain interaction data like ratings or booking counts. Missing interactions are set to zero.

Example of a simplified User-Item Matrix:

	Manali	Himachal	Assam	Goa
U1	5	4	0	1
U2	0	3	0	4
U3	4	0	5	0
U4	0	0	3	2

Step 2 Perform SVD

Decompose RRR into three matrices:

- U: User preferences.
- Σ: Singular values capturing the significance of latent features.
- V^T: Item characteristics.

Mathematically, R=UΣV^T.

Step 3: Truncate the Matrices

Retain only the top kkk singular values to focus on significant patterns, reducing U,Σand V^T to U_k,Σ_k,U_k,V_k^T..

Step 4: Predict Missing Ratings

Reconstruct R as R^=U_kΣ_kV_k^T, estimating ratings for missing interactions. For example, if R^ predicts a rating of 4.3 for "Assam" by U1, it suggests that U1 would likely enjoy visiting Assam.

Step 5: Generate Recommendations

- For each user, recommend top-rated items from R^ that they have not interacted with.

Example Calculation of SVD in Travel Recommendation

 Input Matrix:

- U1 rates "Manali" and "Himachal" highly but hasn’t rated "Assam" or "Goa."

 Output:

- After applying SVD, the system predicts a high score for "Assam" for U1, suggesting it as a suitable recommendation.

Content based Filtering using TF-IDF

Review 1 (Miami): "I loved the beach and the vibrant nightlife."

Review 2 (New York): "The art scene and museums in New York are amazing."

Review 3 (Ibiza): "Ibiza is known for its beaches and nightlife."

Step1:TF(t,d)=Number of times term t appears in documen t d /Total number of words in document d

Review 1 (Miami):Total words = 8

TF(beach, Review 1) = 1/8=0.125 **TF(nightlife, Review 1)** =1/8=0.125

Review 2 (New York): Total words = 9

TF(beach, Review 2) = 0/0=0 **TF(nightlife, Review 2)** = 0/9=0

Review 3 (Ibiza):Total words = 7

TF(beach, Review 3) = 1/7=0.143 **TF(nightlife, Review 3)** = 1/7=0.143

Step 2: Inverse Document Frequency (IDF) IDF(t)=log(Total number of documents/Number of documents containing the term t)

Total number of documents = 3

Documents containing "beach": Review 1 (Miami) and Review 3 (Ibiza)

IDF(beach)=log(3/2)=log(1.5)=0.176

Documents containing "nightlife": Review 1 (Miami) and Review 3 (Ibiza)

IDF(nightlife)=log(3/2)=log(1.5)=0.176

Step 3: TF-IDF Calculation:

TF-IDF(t,d)=TF(t,d)×IDF(t)

TF-IDF for "beach":

Miami: TF-IDF(beach, Miami)=0.125×0.176=0.022

New York: TF-IDF(beach, New York)=0×0.176=0

Ibiza: TF-IDF(beach, Ibiza)=0.143×0.176=0.025

TF-IDF for "nightlife":

Miami: TF-IDF(nightlife, Miami)=0.125×0.176=0.022

New York: TF-IDF(nightlife, New York)=0×0.176=0

Ibiza: TF-IDF(nightlife, Ibiza)=0.143×0.176=0.025

Term	Miami (Review 1)	New York (Review2)	Ibiza (Review 3)
beach	0.022	0	0.025
nightlife	0.022	0	0.025

Fig. 2. Table of TF -IDF

III. EXPERIMENTAL RESULT

Fig 3 provides a scatterplot matrix illustrating relationships among key attributes in the recommendation dataset, including destination rating, price range, and type code. The matrix reveals clustering patterns and correlations that are instrumental for refining recommendation criteria. For instance, destinations are grouped based on similarity in attributes, aiding in the creation of diverse yet relevant recommendations by identifying shared characteristics across various destination types.

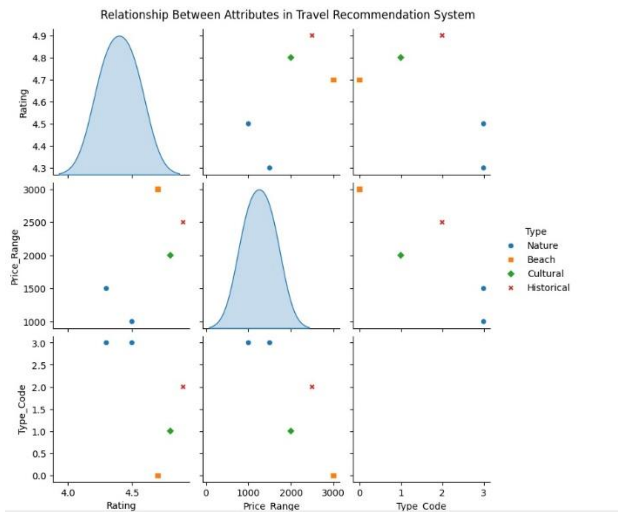

Fig. 3. Attribute Relationships in the Recommendation Dataset

Fig 4 depicts cosine similarity scores between user preferences and available destinations, a critical metric in the collaborative filtering component of the system. High cosine similarity scores indicate a closer match between the user's past preferences and destination attributes, enabling the model to refine recommendations by focusing on options that exhibit higher alignment with user interests.

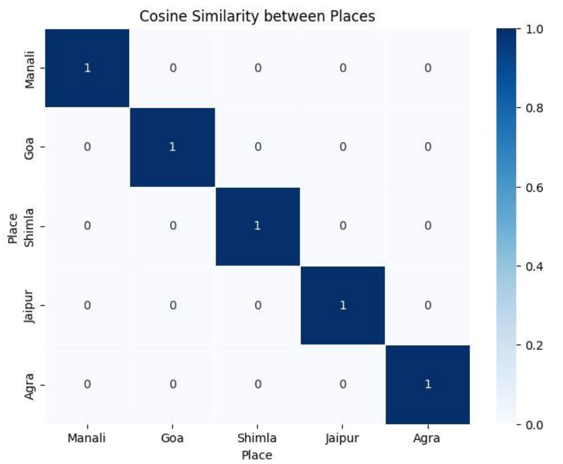

Fig .4. Cosine Similarity Scores Between User Preferences and Destinations

Addressing Cold Start and Data Sparsity

- **Cold Start**: One of the primary challenges in travel recommendation systems is the cold start problem, which occurs when there is insufficient data on new users or destinations. The hybrid system effectively mitigates this issue by using content-based filtering to recommend destinations based on attributes even when interaction data is sparse. The CBF component ensures that new users or destinations receive reasonable initial recommendations, allowing the CF component to improve as more interaction data is collected.

- **Data Sparsity**: The hybrid model also addresses data sparsity by applying matrix factorization (SVD) within the CF component, which reduces dimensionality and uncovers latent patterns in user-item interactions. This approach improves the model's ability to generate recommendations even in cases of limited data availability.

Fig 5 displays the main user interface of the Destinify travel recommendation system. The layout emphasizes a streamlined user experience, with prominent sections for destination search, personalized recommendations, and category-based exploration. The design principles prioritize ease of use, with intuitive navigation and clearly labeled sections. Key functionalities are accessible from the homepage, allowing users to initiate searches and view suggested destinations directly.

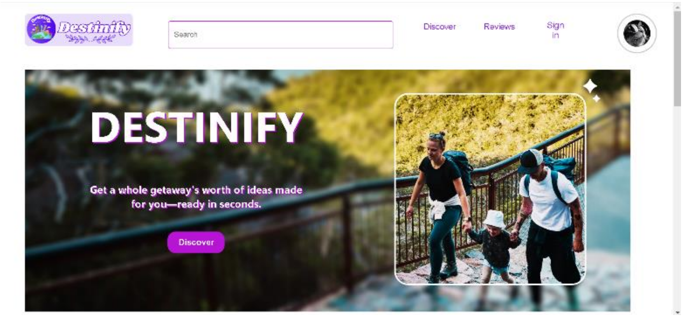

Fig.5 . Destinify Interface

Conclusion

This paper introduced *Destinify*, an intelligent travel recommendation system that leverages both content-based and collaborative filtering methods to deliver personalized travel suggestions. By analyzing user preferences, destination attributes, and historical travel data, Destinify aligns recommendations with individual interests and travel patterns, offering a user-friendly and tailored approach to travel planning. The system demonstrated effective clustering of destinations by similarity and optimized recommendations using cosine similarity scores, providing users with meaningful suggestions based on complex relationships within the data.

Destinify's performance highlights its potential to simplify travel planning with a personalized, data-driven interface. Future developments could enhance its adaptability through real-time feedback, evolving user preferences, and multi-modal data sources, including social media trends and seasonal influences. Overall, Destinify marks an advancement in travel recommendation technology, merging user-centric design with advanced analytics to enrich the modern travel experience.

.

REFERENCES AND RESOURCES

[1] Parivash Pirasteh 1 · Mohamed-Rafik Bouguelia 2 · K. C. Santosh

"Personalized recommendation: an enhanced hybrid collaborative filtering", 25 november 2022.

[2] Abhishek Kulkarni , R. M. Samant , Prathamesh Barve , Aarushi Phade "A Machine Learning Approach to Building a Tourism Recommendation System using Sentiment Analysis" 19 june 2019

[3] Hajare1, Hrushikesh Jarad2, Shabnam Jakate3, Rajkesh Nishad4 And Prof. Mrs. Nilufar Zaman5 ,"TOURS AND TRAVEL RECOMMEND SYSTEM USING MACHINE LEARNING TECHNIQUES",07 july 2022

[4] Choirul Huda, Arief Ramadhan, Agung Trisetyarso, Edi Abdurachman, Yaya Heryadi, "Smart Tourism Recommendation Model",12 december 2022.

S. Bagchi, "Performance and quality assessment of similarity measures in collaborative filtering using mahout," Procedia Computer Science, vol. 50, pp. 229–234, 2015.

[5] Abhishek Kulkarni , R. M. Samant , Prathamesh Barve , Aarushi Phade "A Machine Learning Approach to Building a Tourism Recommendation System using Sentiment Analysis" 19 june 2019

[6] Angulakshmi1 , Rathi.R2 , Dr.Sudha Senthilkumar3 , Dr.K.Brindha4 , Rajat Tandon, "A Study on Travel Recommendation System ", 21 december 2021

[7] Y. Kurata and T. Hara, "Ct-planner4: Toward a more user-friendly interactive day-tour planner," in Information and communication technologies in tourism 2014. Springer, 2013, pp. 73–86

[8] WJ Xiao, SJ Yao and SM Wu, "Improved Top-N collaborative filtering recommendation algorithm [J]", *Application Research of Computers*, vol. 35, no. 1, pp. 1-7, 2017