CONTENT BASED TOURIST PLACE RECOMMENDATION SYSTEM

Abstract:

Tourism is a rapidly growing industry, with travelers seeking personalized and tailored experiences. In response, recommendation systems have emerged as valuable tools to assist travelers in discovering suitable destinations. This paper presents a Content-Based Tourist Place Recommendation System designed to suggest travel destinations based on individual preferences and interests. Unlike traditional approaches that primarily rely on collaborative filtering, our system utilizes content-based filtering to recommend tourist spots. By analyzing features such as location, attractions, the system generates personalized recommendations aligned with user preferences. Text preprocessing techniques, TF-IDF vectorization, and cosine similarity measures are employed to enhance recommendation accuracy. Through this system, travelers can discover new destinations that align with their unique preferences, thereby enhancing their overall tourism experience.

1. INTRODUCTION:

The Content-Based Tourist Place Recommendation System described in the paragraph is a novel approach to assisting travelers in finding destinations tailored to their preferences. Unlike traditional recommendation systems, which often rely on collaborative filtering and user interactions, this system focuses on analyzing the content of tourist spots to generate personalized recommendations. By examining various features such as location and attractions, the system aims to understand the unique preferences and interests of individual travelers. This content-based approach allows for more targeted and relevant recommendations, ensuring that users are presented with destinations that align closely with their desires.

In terms of features analyzed, the system considers several key aspects of tourist spots to make recommendations. Location plays a crucial role, as travelers often have preferences regarding the geographic region they wish to visit. Additionally, the attractions available at each destination are carefully analyzed to match users with spots that offer activities and experiences they are likely to enjoy. Other features such as cultural significance, historical landmarks, and natural beauty may also be taken into account to provide a comprehensive recommendation tailored to each user's preferences.

To achieve accurate recommendations, the system employs several text preprocessing techniques and algorithms. Text preprocessing involves cleaning and transforming textual data to extract meaningful information. Techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) vectorization are used to represent the textual features of tourist spots numerically. Cosine similarity measures are then applied to compare the preferences of users with the characteristics of different destinations, enabling the system to generate recommendations that closely match individual interests. Overall, this content-based approach enhances the tourism experience by offering personalized recommendations that cater to the unique preferences of each traveler.

2. LITERATURE SURVEY:

SNo	Title	Authors	Summary	Drawbacks
1	A Multi-Level Tourism Destination Recommender System (2020)	Hend Alrasheed , Arwa Alzeer, Arwa Alhowimel, Nora shameri, Aisha Althyabi	The paper proposes a multi-level tourism destination recommender system to assist potential travelers in finding destinations that match their preferences and requirements. The system incorporates two procedures: providing the user with a set of destinations liked by similar users and ranking the destinations based on user preferences and constraints. The system utilizes user preferences, including attraction types and weather, along with constraint attributes like travel dates and budget.	• Lack of Empirical Evaluation • Lack of Comparative Analysis
2	Hybrid Recommender System for Tourism Based on Big Data and AI: A Conceptual Framework (2021)	Khalid AL Fararni, Fouad Nafis, Badraddine Aghoutane, Ali Yahyaouy, Jamal Riffi, Abdelouahed Sabri	The paper proposes a conceptual framework for a hybrid recommender system in the field of tourism, specifically targeting the Daraa-Tafilalet region in Morocco. The ultimate goal is to design a recommender system based on big data technologies, artificial intelligence, and operational research to enhance tourism in the specified region. The proposed system not only recommends tourist attractions but also acts as a trip planner, creating detailed itineraries based on user preferences.	• Limited Evaluation Metrics • Generalization to Other Regions

3	Tour Spot Recommendation System Via Content-Based Filtering (2022)	Mishal Muneer, Uzair Rasheed, Sadia Khalid, Muhammad Ahmad	The paper proposes a Tour Spot Recommendation System through content-based filtering to enhance the tourist experience by recommending the best picnic spots based on user preferences, budget, and interests. The system utilizes a dataset containing information about various tour spots, including features such as country, region, geography, budget, safety, transport, climate, and descriptions. The recommendation system employs text preprocessing techniques, TF-IDF for feature weighting, and cosine similarity for finding similar tour spots. Two recommendation engines are designed, one for existing users based on their previous history and another for new users who provide their preferences through a form.	• Relying on user provided data • Does not address dynamic factors such as seasonal variation • Evaluation Metrics
4	Design and Implementation of a Personalized Tourism Recommendation System Based on the Data Mining and Collaborative Filtering Algorithm(2022)	Xiang Nan, Kayo kanato, Xiaolan Wang	The paper introduces a Collaborative Mining and Filtering Process (CMFP) for a Personalized Tourism Recommendation System. The proposed system aims to enhance recommendation efficiency and data analysis by utilizing data mining and collaborative filtering algorithms. The system considers various contextual factors such as social media sentiment, weather, user preferences, time, and location to provide more accurate tourism recommendations. The proposed system is analyzed using metrics such as accuracy, data handling rate, mining time, and overhead.	• Lack of Comparative Analysis

| 5 | Personalized Travel Recommendation Systems: A Study of Machine Learning Approaches in Tourism (2023) | Mohamed Badouch, Mehdi Boutaounte | This paper provides an extensive overview of recommender systems (RS) in the tourism sector, covering different recommendation approaches, stages of travel, data sources, and phases of data processing. It discusses three main types of RS - content-based, collaborative filtering, and hybrid systems - highlighting their strengths and weaknesses. The paper identifies three stages of travel: before, during, and after, emphasizing the role of RS in trip planning, real-time decision support, and post-trip feedback analysis. It concludes by underlining the importance of RS in enhancing tourism experiences and the need for continued research and development in the field. | Evaluation complexity Limited interpretability Complexity in design and implementation |

3.1 PROBLEM STATEMENT:

Developing a content-based tourist place recommendation system presents the challenge of integrating diverse destination attributes and user preferences to offer personalized travel suggestions. This entails collecting and preprocessing comprehensive data on tourist destinations, understanding user preferences encompassing demographics, past travel history, and specific requirements, and designing an algorithm that efficiently matches user profiles with suitable destinations. Ensuring scalability and real-time performance, alongside establishing evaluation and feedback mechanisms, are critical for enhancing recommendation accuracy and user satisfaction, ultimately contributing to the advancement of personalized travel experiences worldwide.

3.2 PROPOSED FRAMEWORK:

In the devised methodology, we leveraged content-based filtering to suggest optimal picnic spots to users. This approach entails recommending picnic spots based on their similarity, considering both the characteristics of the spots and the user's past preferences and then recommending the nearby hotels and restaurants for those recommended picnic spots. Implementation of the recommendation system was conducted using Python programming language. Initially, a comprehensive dataset comprising all tour spots, restaurants and hotels was constructed, encompassing attributes detailed in Table 1, Table 2,Table 3. Subsequently, text data underwent preprocessing procedures, followed by conversion into vectors utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) technique. Through similarity analysis, the system

identifies and suggests the most suitable picnic spot to the user. The process flow is illustrated in Figure 1.
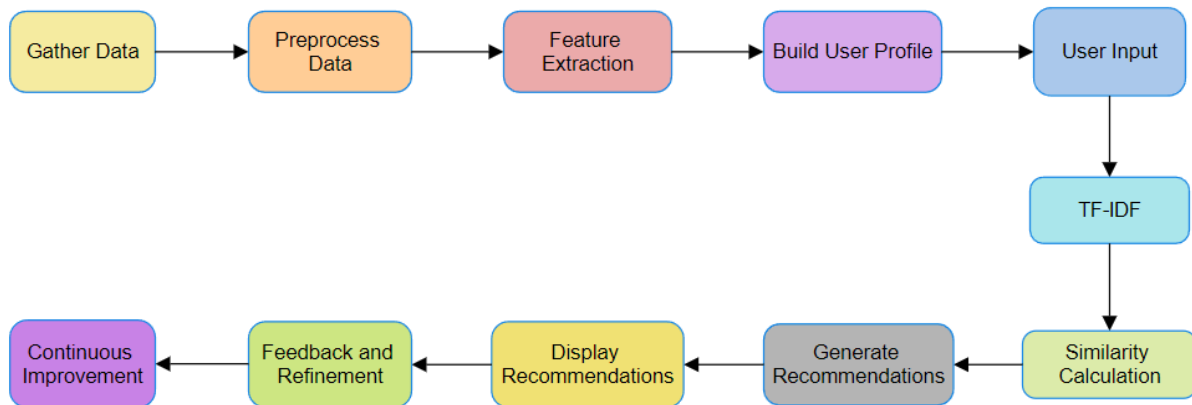

Figure 1. Overview of methodology

4.METHODOLOGY USED:

A. Dataset:

Three meticulously curated datasets harvested from TripAdvisor via the powerful automation tool, Instant Data Scraper, form the backbone of our content-based recommendation system. These datasets are rich repositories of information encompassing attractions, hotels, and restaurants. Each dataset has been meticulously collected to ensure comprehensive coverage and accuracy, empowering our recommendation engine to deliver personalized and insightful recommendations to users.

Table1. Overview of features in Attractions dataset

Features	Values	Description
Place_url	Urls to places in tripadvisor	This column has the url's to the attractions. These urls are directed to tripadvisor.
AttractionImage_url	Urls to images of the place	This column has the urls to images of the place. These urls are directed to tripadvisor.
Name Of The Place	Places names	This column has names of different places in Hyderabad, having tour spots.

URL_TO_REVIEWS_PAGE	Urls of customer review page in tripadvisor	This column has the url's to customer reviews page which are directed to the tripadvisor page.
Rating	1-5	This column have the overall rating given by the people who visited the place
No_of_reviews	Total no of reviews	This column have the count of no of reviews given by people
Themes	Amusement,theme parks,religious sites,zoos,bodies of water,gardens, shopping malls, spas, parks, national parks, water parks, games, Entertainment centers	This column has themes of the places as the value.
Address	Address of the places	This column has the addresses of the places.
Description	Textual Description	This column has the description about the attractions.

Table2. Overview of features in Restaurants dataset

Features	Values	Description
link_to_swiggy	Urls to swiggy	This column has the url's to the restaurants. These urls are directed to swiggy.
images	Urls to images	This column has the url's to the images of restaurants. These urls are directed to tripadvisor.
Restaurant_names	Names of the restaurants	This column has names of restaurants in Hyderabad.

Rating	1-5	This column has the overall rating given by the people who visited the restaurant.
Restaurant Type	Themes of the restaurants	This column has the type of dishes or food restaurants serve.
Location	Locations of the places	This column has the locations of the places.
Descriptions	Textual Description	This column has the description about the restaurants.

Table3. Overview of features in hotels dataset

Features	Values	Description
URL_TO_HOTEL	Urls to places in MakeMyTrip	This column has the url's to the restaurants. These url's are directed to MakeMyTrip.
URL_TO_HOTEL_IMAGE	Urls to images	This column has urls to the images of hotels. These url's are directed to MakeMyTrip.
Hotel_Name	Names of the hotels	This column has names of hotels in Hyderabad.
Area	Locations of the hotels	This column has the area in which hotels are located.
Suitable_For	Couple Friendly, Suitable for anyone.	This column has the details about to whom the hotel is suitable.
Description_about_Hotel	Textual description.	This column has the description about the hotels.
Address_Of_The_Hotel	Location to the hotel.	This column has the exact address of the hotel.
Facilities	List of Facilities available in the hotels	This column contains the facilities which are made available to the customers by

		the hotels.
ratingText	Excellent, Very Good,Good	This column has the overall rating of the hotels in the form of text.
Rating	1-5	This column has the overall rating given by the people who visited the hotel.
No_Of_Ratings	Count of the rating received	This column has the total count of the number of ratings received from the customers.
Actual_Price_per_night	Actual cost to be paid per head per night	This column has the actual price the customer needs to pay per head per night stay in the hotel.
Price_With_Discount	Price after excluded the discount amount	This column has the amount that the customer has to pay after excluding the discount amount from the actual price.
Tax	Additional tax amount to be paid	This column includes the additional tax amount that the customer need to pay.
Offers	Offers available	This column contains the additional offers which are made available to the customers by the hotels.
Cancellation_policy	No cancellation policy / Free Cancellation policy until specified time.	This column contains the cancellation policy of the hotels.

B. Preprocessing:

Handling missing values is a critical step in data preprocessing to ensure that the data used for analysis or modeling is accurate and reliable. We used several techniques for handling missing values in Python, which are outlined below:

- Identifying Missing Values: Before handling missing values, it's essential to identify where they exist in the dataset. In Python, missing values may be represented as NaN (Not a Number) in NumPy or pandas, or as None in Python objects.

- Removing Missing Values: If the missing values are relatively few and randomly distributed and are of less importance, we might choose to remove rows or columns containing missing values.
- Imputation: Imputation involves replacing missing values with substituted values. This can be done using various statistical measures such as mean, median, mode, or using more sophisticated methods like predictive modeling.
- Using a Placeholder: Sometimes, missing values can be replaced with a specific placeholder value to indicate their absence.

C. Feature Extraction:

Feature extraction is particularly useful when dealing with high-dimensional data or when certain features are noisy, irrelevant, or redundant.

- The process of feature extraction in a dataset involves the identification and removal of irrelevant or redundant attributes, while retaining only the most relevant and important features for analysis.
- In Excel, this process often entails manually reviewing the dataset and selecting the columns (features) that are deemed essential based on domain knowledge, statistical significance, or the objectives of the analysis.
- Unwanted attributes, such as irrelevant identifiers, redundant variables, or those with high levels of missing data, are systematically removed to create a more concise and focused dataset.

D. WORKING OF THE CODE:

Dataloading and Preprocessing:

- The script begins by loading data from Excel files containing information about visited places (user_visited_df) and all places (all_places_df). The 'DESCRIPTION' column in the all_places_df dataframe is filled with empty strings where missing.
- It utilizes the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization technique to convert text descriptions of places into numerical vectors, which are then used to calculate the similarity between visited places and all places.

Recommendation Function:

- This function takes the user's visited places, all places data, cosine similarity matrix, and an optional parameter top_n (default set to 5) to recommend places.
- It iterates over each visited place, finds similar places based on cosine similarity scores, and filters out recommendations that are not contextually relevant to the visited place's category.
- The function returns a dataframe containing recommendations.

Recommendation Generation:
- After defining the recommendation function, it recommends places to the user, stores the recommendations in an Excel file, and prompts the user to select a place from the recommendations.

Fetching Additional Information:
- Once the user selects a place, the script fetches the location of the selected place and proceeds to recommend top hotels and restaurants in that location.

Fetching Top Hotels and Restaurants:
- Two functions, get_top_hotels and get_top_restaurants, retrieve top hotels and restaurants respectively based on the target location provided by the user.
- These functions load hotel and restaurant data from CSV and Excel files respectively, filter them by the target location, sort them by ratings, and return the top entries.

Displaying Recommendations:
- Finally, the script displays the recommended top hotels and restaurants for the selected place.

5. RESULTS:

The recommendation system operates in two stages: firstly, by analyzing the user's browsing history to suggest similar spots based on past preferences, and secondly, by utilizing the selected place's location to recommend top restaurants and hotels nearby. This personalized approach ensures that users receive tailored recommendations aligned with their interests and facilitates a seamless exploration experience, combining user preferences with location-based insights for enhanced satisfaction.After user selects a place the script fetches the location of the selected place and it recommends top hotels and restaurants.

Let's consider the below example, there are list of places visited by the user previously which can be seen in the left side and the right side consists of the corresponding recommended places for the users previously visited places.

```
           Visited Place                        Recommended Place  \
0          Ramoji Film City                      Ramoji Film City
1            Golconda Fort                         Golconda Fort
2            Golconda Fort                       Rachakonda Fort
3        Wonderla Hyderabad                    Wonderla Hyderabad
4        Wonderla Hyderabad                           Thrill City
5        Wonderla Hyderabad       Mount Opera Multi-Theme Park Resort
6              Birla Mandir                          Birla Mandir
7              Birla Mandir       Sri Venkateshwara Swamy Temple
8              Birla Mandir                          Shyam Mandir
9              Birla Mandir  ISKCON Hyderabad Sri Sri Radha Madanmohan Mandir
10       Salar Jung Museum                     Salar Jung Museum
11    Taj Falaknuma Palace                  Taj Falaknuma Palace
12    Taj Falaknuma Palace                   Chowmahalla Palace
13               Charminar                             Charminar
14   Chilkur Balaji Temple                 Chilkur Balaji Temple
15   Chilkur Balaji Temple             Balkampet Yellamma Temple
16   Chilkur Balaji Temple        Sri Venkateshwara Swamy Temple
17   Chilkur Balaji Temple                        Sanghi Temple
18      Chowmahalla Palace                    Chowmahalla Palace
19      Chowmahalla Palace                    Chowmahalla Palace
20      Chowmahalla Palace                  Taj Falaknuma Palace
21           Calvary Temple                        Calvary Temple
22   Shri Jagannath Temple                 Shri Jagannath Temple
23   Shri Jagannath Temple                 Nagendra Swamy Temple
```

For example Nagendra Swamy Temple is the place that was selected by the user from the above recommended places.
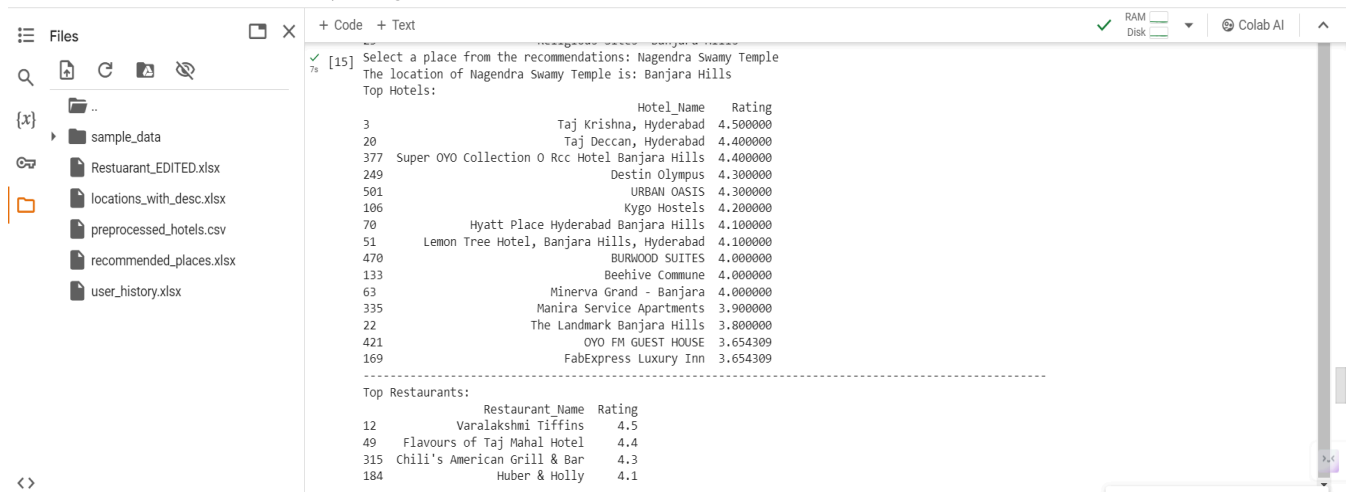
+ Code   + Text

```
22   Shri Jagannath Temple                    Shri Jagannath Temple
23   Shri Jagannath Temple                    Nagendra Swamy Temple

                              Category   Location/Area
0          Amusement , Theme Parks    Sultan Bazar
1          Historic Sites , Castles   Sultan Bazar
2          Historic Sites , Castles            NaN
3          Amusement , Theme Parks        Open now
4          Amusement , Theme Parks        Open now
5          Amusement , Theme Parks    Sultan Bazar
6                   Religious Sites    Sultan Bazar
7                   Religious Sites             NaN
8                   Religious Sites             NaN
9                   Religious Sites           Abids
10                    Art Museums    Sultan Bazar
11          Architectural Buildings   Sultan Bazar
12          Architectural Buildings      Charminar
13  Religious Sites , Monuments , Statues   Sultan Bazar
14                  Religious Sites             NaN
15                  Religious Sites             NaN
16                  Religious Sites             NaN
17                  Religious Sites             NaN
18          Architectural Buildings      Charminar
19          Architectural Buildings             NaN
20          Architectural Buildings   Sultan Bazar
21                  Religious Sites        Open now
22                  Religious Sites   Banjara Hills
23                  Religious Sites   Banjara Hills
Select a place from the recommendations: Nagendra Swamy Temple
```

```
                    Religious Suites Banjara Hills
[15]  Select a place from the recommendations: Nagendra Swamy Temple
  7s  The location of Nagendra Swamy Temple is: Banjara Hills
      Top Hotels:
                                            Hotel_Name    Rating
      3                             Taj Krishna, Hyderabad  4.500000
      20                             Taj Deccan, Hyderabad  4.400000
      377  Super OYO Collection O Rcc Hotel Banjara Hills  4.400000
      249                                 Destin Olympus  4.300000
      501                                    URBAN OASIS  4.300000
      106                                   Kygo Hostels  4.200000
      70            Hyatt Place Hyderabad Banjara Hills  4.100000
      51    Lemon Tree Hotel, Banjara Hills, Hyderabad  4.100000
      470                                 BURWOOD SUITES  4.000000
      133                                Beehive Commune  4.000000
      63               Minerva Grand - Banjara  4.000000
      335                     Manira Service Apartments  3.900000
      22              The Landmark Banjara Hills  3.800000
      421                            OYO FM GUEST HOUSE  3.654309
      169                          FabExpress Luxury Inn  3.654309
      -----------------------------------------------------------------
      Top Restaurants:
                         Restaurant_Name  Rating
      12             Varalakshmi Tiffins     4.5
      49        Flavours of Taj Mahal Hotel     4.4
      315    Chili's American Grill & Bar     4.3
      184                  Huber & Holly     4.1
```

After User gives the place from the recommended places the script fetches the top Restaurants and Hotels to the user based on the location of that particular place as shown in the above picture.

6. CONCLUSION AND FUTURE WORK:

The development of our content-based tourist recommendation system marks a significant advancement in enhancing user experience and aiding informed decision-making in the tourism industry. Throughout the project, we meticulously explored various facets of content-based recommendation systems, encompassing data collection, preprocessing, feature extraction, and recommendation generation. By harnessing natural language processing and sentiment analysis, we effectively extracted meaningful insights from textual descriptions of tourist attractions, enabling us to capture the essence of each destination accurately.

We would like to extend our project by integrating it with google maps api for more personalized recommendations according to the user location.

References:

1. A. Kontogianni, E. Alepis, and C. Patsakis, "Promoting smart tourism personalized services via a combination of deep learning techniques," *Expert Systems with Applications*, vol. 187, Article ID 115964, 2022.

   View at: Publisher Site | Google Scholar

2. F. Santos, A. Almeida, C. Martins, R. Gonçalves, and J. Martins, "Using POI functionality and accessibility levels for delivering personalized tourism recommendations," *Computers, Environment and Urban Systems*, vol. 77, Article ID 101173, 2019.

   View at: Publisher Site | Google Scholar

3. C. Bin, T. Gu, Y. Sun, and L. Chang, "A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining," *Multimedia Tools and Applications*, vol. 78, no. 24, pp. 35135–35156, 2019.

   View at: Publisher Site | Google Scholar

4. S. Missaoui, F. Kassem, M. Viviani, A. Agostini, R. Faiz, and G. Pasi, "LOOKER: a mobile, personalized recommender system in the tourism domain based on social media user-generated content," *Personal and Ubiquitous Computing*, vol. 23, no. 2, pp. 181–197, 2019.

   View at: Publisher Site | Google Scholar

5. S. Du, H. Zhang, H. Xu, J. Yang, and O. Tu, "To make the travel healthier: a new tourism personalized route recommendation algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 9, pp. 3551–3562, 2019.

   View at: Publisher Site | Google Scholar