A tag-based recommender system for tourism using collaborative filtering

Afef Selmi, Maryah Alawadh, Raghad Alotaibi, Shrefah Alharbi
Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

Article Info

Article history:

Received Apr 18, 2024
Revised Oct 30, 2024
Accepted Nov 11, 2024

Keywords:

Collaborative filtering
Recommendation system
Tag module
Tailored suggestions
Tourism industry
User preferences

ABSTRACT

Recommender systems have garnered significant attention from researchers due to their potential for delivering personalized recommendations in light of the vast amount of information available online. These systems have found applications in various domains, including financial services, movies, and research articles. Their implementation in the tourism industry is particularly promising. Travelers often face the daunting task of selecting the right tourist attractions from a plethora of options, which can consume considerable time and energy. By leveraging personalized recommendation technologies, it is possible to provide highly accurate travel suggestions tailored to individual preferences. This study proposes the development of a customized recommendation system (RS) aimed at assisting travelers in the Qassim region of the Kingdom of Saudi Arabia. By using this region as a case study, the proposed RS consists of two main modules: a user registration and login module and a recommendation technique and tag module. The system will capture users' interests and prompt them to select from various options, subsequently presenting them with tailored recommendations based on their preferences. This approach aims to enhance the travel experience by offering relevant suggestions that align with the interests of each traveler.

Corresponding Author:

Afef Selmi
Department of Information Technology, College of Computer, Qassim University
Buraydah 51452, Saudi Arabia
Email: a.selmi@qu.edu.sa

1. INTRODUCTION

Traveling has always been an essential aspect of human existence, serving various purposes throughout history. Tourism, in particular, stands out as one of the most significant sources of income for nations, reflecting their culture and civilization. This study will explore how modern recommendation systems (RS) have positively impacted the travel and tourism industry. In recent years, the tourism sector has experienced remarkable growth, fueled by advancements in technology and changes in consumer behavior. As travelers seek personalized experiences, RS have emerged as powerful tools to enhance their journeys. These systems analyze vast amounts of data, including user preferences, historical behavior, and real-time information, to provide tailored suggestions for attractions, accommodations, and activities. The integration of contemporary RS into the travel and tourism industry has revolutionized how travelers plan and experience their journeys. As the industry continues to grow, these systems will play a crucial role in shaping future travel experiences, making them more personalized and enjoyable.

The rapid worldwide progress in information technology and communications, along with the widespread use of the Internet, contributed to this growth. It enabled easier access to a wealth of global data on tourist destinations, travel itineraries, and points of interest (POI) for both customers and travelers. The e-tourism industry has grown dramatically. There are numerous businesses that provide travel packages.

But frequently, they aren't appropriate. Large-scale tourist information and plans also contribute to the general dissatisfaction and dispersal of travelers. The availability of traveler recommendations will be very helpful in determining which location is best for each individual. It also helps users make the right decisions by considering their emotions and unique interests. With the rise of information on the internet, RS have proven to be an effective solution for addressing information overload across various fields [1], [2]. The use of RS has gained attraction as a research topic in recent years because of its many benefits in terms of appropriate recommendation-making.

Recommender systems facilitate the provision of tailored suggestions or the individual user's guided navigation to interesting or helpful items within a significant data space [3], including travel [4], cloud services [5], interest of things (IoT) [6], books [7], and e-learning [8]. Many recommendation strategies have been put forth in an effort to accomplish this [9]. The aim of this paper is to present an overview of popular recommendation techniques. In particular, it examines three widely recognized methods: content-based, collaborative filtering-based, and hybrid approaches [1], [2]. Lastly, it provides some key challenges with strong justifications for this emerging and growing field, specifically in tourism.

Travel is a vast sea of information, and developing an effective tourism strategy demands time, energy, and expertise. The personal tourism RS for Saudi Arabia's Qassim region aims to streamline this process by suggesting optimal travel itineraries. This system adjusts to the interests and preferences of tourists and visitors, efficiently gathering, analyzing, and presenting information to craft personalized itineraries.

This study investigates the effects of personalized RS on enhancing tourist experiences in the Qassim region of Saudi Arabia. While previous studies have explored the impact of RS in tourism, they have primarily focused on general models without addressing the specific influence of local cultural and geographic factors on tourists' preferences and experiences. Additionally, earlier research has not explicitly examined how these systems can dynamically adjust to changing user preferences within a specific region. This study fills these gaps by tailoring the RS to the unique characteristics of the Qassim region, aiming to provide a more targeted and culturally relevant travel experience.

This study introduces a personalized RS specifically designed for the Qassim region. Its main objective is to provide tailored recommendations and suggestions based on user preferences and stored data. The tourism RS utilizes filtering techniques to identify and recommend destinations that promise an ideal experience. A key advantage is that travelers can customize their itineraries, significantly reducing the time and effort required to find locations that match their interests.

– Motivation

a. Proposing a personalized RS for Qassim region of the Kingdom of Saudi Arabia: the proposed system will help to provide pertinent recommendation. It also alters visitors' and tourists' interests and inclinations. To help the user design a suitable itinerary, the personal tourism RS collects, assesses, and displays information in the most efficient way possible.
b. Saving time and effort: the proposed system is designed to offer recommendations and suggestions based on user preferences and collected data. The system suggests travel destinations that offer the best possible experience, allowing tourists to save time and effort by customizing their schedules to match their preferences and avoiding locations that don't align with their interests.

– Highlights

The main objectives of the present work include:

a. Introducing an overview of the concepts of RS, the tourism RS and the most used literature techniques of RS and discussed their strengths and weaknesses.
b. Reviewing the current recommendation-based approaches focused on tourism field. We discuss the strengths and the weaknesses of these proposed approaches to identify the appropriate technology for our proposal system.
c. Proposing a personalized tourism RS for Qassim region of the Kingdom of Saudi Arabia using collaborative filtering as a recommendation technique.
d. Assessing the effectiveness of the suggested RS by utilizing a real dataset gathered from various sources, including Google Maps. The dataset contains Qassim region data regarding user information, attractions, chalets, heritage landmarks, parks, restaurants, public places, malls, and hotels.

The structure of this paper is outlined as follows: section 2 provides an overview of RS. Section 3 surveys the current literature on the subject. Section 4 details the research methodology employed in this study. Section 5 addresses the system's evaluation, encompassing data collection, analysis, and interpretation. Lastly, section 6 summarizes the paper and discusses potential future research directions. Figure 1 illustrates the paper's roadmap.

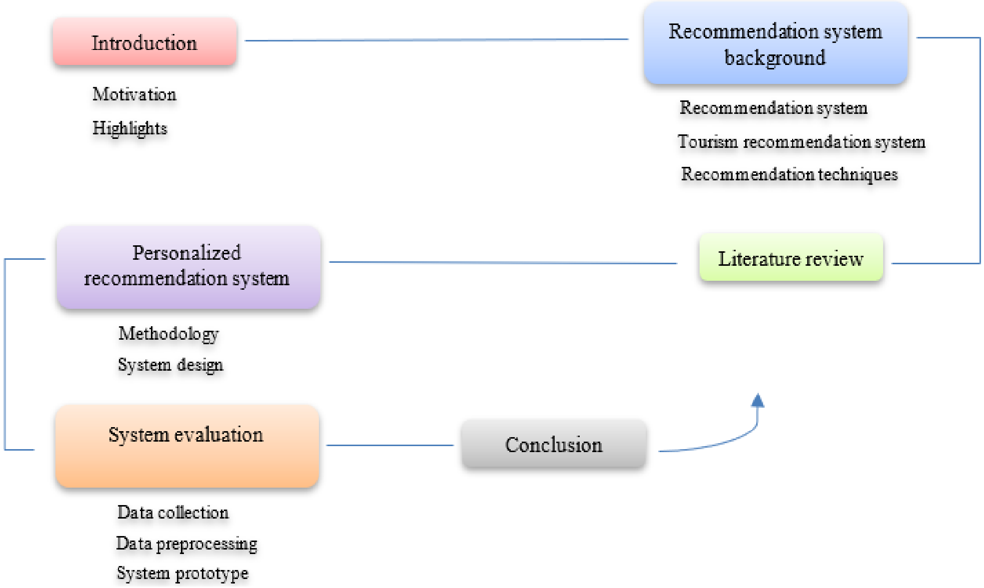

Figure 1. Paper roadmap

2. RECOMMENDATION SYSTEM: BACKGROUND

This section briefly introduces the concepts of RS, the tourism RS and the most used literature techniques of RS.

2.1. Recommendation system

RS are tools that interact with large information spaces. They aim to reduce the search effort of users by providing them meaningful recommendations or by directing them to useful resources within a large data space [2], [3]. RS leverage the opinions of community members to help individuals find information or products that are most likely to be of interest or relevance to them. Today, RS is implemented on nearly all e-commerce websites, aiding a vast number of users. These systems typically utilize collaborative filtering, content-based filtering (also referred to as the personality-based approach), or a combination of both, known as hybrid-based filtering, along with other systems like knowledge-based systems.

2.2. Tourism recommendation system

A tourism RS serves as a personalized tool for tourists, helping them efficiently search for attractions. It is regarded as an effective way for tourists to discover places of interest [10], [11]. The system compares collected data with similar and different data from other users, generating a list of recommended attractions for the tourist. In the tourism industry, modern technologies from traditional recommender systems, such as collaborative filtering, are considered successful. First, the system offers a list of city attractions likely to appeal to the user, considering their demographic profile, past preferences, and current visit interests. Second, a planning module organizes the suggested venues by considering their timing and the user's constraints, determining how and when to enjoy the recommended activities. A key feature missing in most recommender systems is the ability to present the recommended actions in the form of a structured agenda, or an actionable plan.

2.3. Recommendation techniques

2.3.1. Content-based filtering systems

Content-based filtering, also referred to as cognitive filtering, is one of the most popular and extensively studied recommendation methods. A crucial aspect of this technique is the user modeling process, where a user's interests are inferred based on the items they engage with. This involves comparing the content of an item with the user's profile [2].

A content-based filtering system utilizes the term frequency (TF) and inverse document frequency (IDF) method [2], [7]. This approach gathers data by calculating both the frequency of a term and its inverse document frequency. The goal is to assess the significance of a word in a text by counting how often it appears within a document and determining how many documents contain that word. Frequently occurring words are assigned a higher score. The process can be summarized as follows: i) TF is the number of times a term appears in a document divided by the total number of terms, ii) IDF is the logarithm of the total number of documents divided by the number of documents containing the term, and iii) the final score is the product of TF(t) and IDF(t).

2.3.2. Collaborative filtering systems

Collaborative filtering, also referred to as social filtering, generates recommendations based on the preferences of users with similar tastes [1]. It makes suggestions by analyzing a user's transaction history, ratings, or purchase data. A distinctive feature of this technique is that it does not take the item's attributes into account [2]. Collaborative filtering systems typically employ two main techniques:

- User-based filtering: this method evaluates how closely aligned a user's preferences are with those of other users. It identifies users with similar characteristics to the target user and then recommends items based on what those similar users like and how they have rated various items.
- Item-based filtering: similarly to user-based filtering, this method measures the similarity between the target item and other items. It analyzes the user-item matrix, identifies items rated by active users, and then compares those items to the target item to generate recommendations.

2.3.3. Knowledge-based systems

Knowledge-based systems rely entirely on domain-specific knowledge. They identify the user's needs and preferences, generating suitable recommendations based on their knowledge models. To address the cold-start problem, knowledge-based systems are often suggested as an alternative to content-based and collaborative systems. However, this approach is computationally intensive, as it requires the acquisition and representation of knowledge for all items in the recommender system [12].

2.3.4. Demographic systems

In demographic filtering, users are grouped based on demographic characteristics to generate recommendations. Many companies use this method for providing recommendation services because it is simple, easy to implement, and computationally efficient. The demographic approach requires market research to uncover relationships between different user groups. The most common way to analyze these connections is by examining user profiles from a company's database or conducting surveys among users. Unlike content-based or collaborative systems, this method does not rely on previous user transaction data, which is a key advantage [12].

2.3.5. Hybrid-based filtering systems

A hybrid filtering strategy integrates multiple filtering methods. It is designed to address common issues associated with content-based filtering, collaborative filtering, and knowledge-based techniques, such as the cold-start problem, over-specialization, and data sparsity. Hybrid filtering is introduced to overcome these challenges and enhance both the accuracy and efficiency of the recommendation process [2]. The hybrid-based filtering system incorporates the following techniques:

- Weighted: this method combines the scores from different recommendation components statistically, aggregating them using an additive formula.
- Switching: the system selects a specific recommendation component from the available options and applies it.
- Feature combination: in this approach, two distinct components contribute to the recommendation process. The actual recommender operates based on data modified by the contributing component, which transfers features from one source to another.
- Feature augmentation: similar to feature combination, but the contributing component provides new features. This method is more flexible than feature combination.
- Cascade: this technique serves as a tie-breaker. Each recommender is assigned a priority, and lower-priority recommenders act as tie-breakers for higher-priority ones.

3. RELATED WORK

Several studies on RS in the tourism domain have been introduced in the literature. This section offers an in-depth review of current research and proposed methods. We analyze the strengths and weaknesses of these approaches to determine the most suitable technology for our proposal.

In their study Lee *et al.* [4], proposed a personalized RS that considers user preferences, diversity, and distance. The system recommends tourist attractions based on three scores: p-Div, p-Pop, and p-Dis. To evaluate the approach, the authors collected user ratings and metadata of POIs from TripAdvisor and Naver. The results demonstrated that the proposed method outperformed existing ones, showing improvements in precision by 2%, 24%, and 20% and recall by 129%, 35%, and 22% for top-1, top-2, and top-3 recommendations, respectively. Additionally, considering users' varied preferences proved effective in enhancing recommendations.

The internet offers a vast amount of information on tourism and leisure activities, making it challenging for travelers to sift through it all to plan their trips. In the study [12], proposed two solutions: i) by applying big data concepts and leveraging data from online communities, contextualization can be broadened to cover multiple dimensions for travelers, addressing various aspects of their journey; and ii) developing new recommender systems as personalization and contextualization continue to expand. Current systems can generate highly personalized, real-time, and context-aware recommendations.

Abbasi-Moud *et al*. [13] proposed a tourism RS that combines semantic clustering and sentiment analysis. The system utilizes contextual information, including the user's location (to suggest nearby attractions), time (to determine when attractions can be visited), and weather (to assess the feasibility of visiting attractions), to provide more suitable recommendations. The goal of this approach is to enhance the recommendation process by incorporating all relevant contextual factors. In addition to improving user satisfaction and convenience, sentiment analysis is also employed to offer a context-aware RS. However, a key drawback of these approaches, which rely solely on location as context information, is their one-dimensional nature and domain limitations.

Using personalized recommendation technology, the authors [14] developed a framework to boost customer loyalty and enhance the flow of tourism products between online and offline platforms. The framework includes a design for personalized online tourism recommendations, covering the process from data collection to the selection of a personalized recommendation algorithm, and integrating content-based recommendations. A single-recommendation method was adopted as the primary approach to increase customer engagement and improve the efficiency of transitioning tourism products from online to offline. Additionally, the framework offers a recommendation strategy for the future operation and management of online tourism projects.

Recently, RS have gained significant attention. Studies show that many existing tourism RS provide misleading suggestions that fail to meet tourists' expectations. A key reason for this issue is the lack of consideration for previous user reviews. To address this, the study [15] proposes a system that incorporates user reviews into the recommendation process. The proposed system is based on three factors: the number of reviews, ratings, and sentiment. User reviews are analyzed and used to recommend hotels to tourists, utilizing a dataset of European hotels. The system's architecture includes four components: input information, user reviews (topics), recommendation techniques, and output information. The system considers the location and user reviews, applying content-based recommendations to present the ideal hotel to the user in both textual and graphical formats.

In this study [16], presented the current progress of a thesis project focused on developing new perspectives for recommender systems in the tourism domain, with a particular focus on POI recommender systems. The project incorporates additional contexts, such as geographical and temporal information, to enhance and evaluate the recommendations. The authors first analyzed the differences between POI RS and traditional RS, emphasizing the importance of contextual information, especially geographical influence, and the challenge of data sparsity. They also discussed various strategies used for both POI and sequential recommendations. Additionally, the authors proposed extending traditional information retrieval metrics to measure recommendation quality, considering dimensions beyond relevance, such as the temporal and geographical context, and information related to users and items to better assess the recommendations' usefulness for the user.

This study [17] introduced a web agent-based intelligent recommendation application for the smart tourism sector, which combines real-time data with a hybrid filtering system to suggest tour packages tailored to client requests. The authors presented an online autonomous web agency designed for the smart tourism ecosystem, built on the integration of various sectors within the tourism industry. They proposed an effective filtering system to enable package customization in the tourism sector.

The study [18] addressed the challenge users face in quickly locating information of interest. To tackle this issue, the authors combined a genetic algorithm (GA) with a long short-term memory (LSTM) network to predict the popularity and ratings of tourism services. They considered the multidimensional social information of users within social networks. Experimental results indicated that this approach significantly outperformed other recent recommendation methods in terms of recommendation quality.

In this study [19], the authors aimed to tackle the cold-start problem in RS. They proposed a hybrid filtering approach that combines content-based filtering, collaborative filtering, and demographic filtering to address this issue. To predict ratings for new users and identify similar items within a neighborhood, the hybrid filtering method utilizes demographic information. The proposed system relies on two algorithms: new user cold-start and new package cold-start. These algorithms extract demographic information such as age and country, age and city, or age and gender, based on various combinations of demographic details. The system then recommends tourism packages that align with the newly predicted ratings for new users using collaborative filtering. The results indicated that employing hybrid filtering alongside demographic filtering is effective in addressing the cold-start problem for new items or new POIs. However, it was noted that many datasets lack demographic details, which are crucial for the recommendation process.

The challenge addressed in the study [20] was the issue of information overload, as users faced difficulty sifting through vast amounts of data online, which consumed considerable time. To mitigate this, the authors proposed a RS that employs tags and collaborative filtering. The system operates in two main steps. The first step involves constructing a model that captures tourists' interests in various attractions. Tags associated with tourist attractions serve as features for this preference model, and a multi-dimensional relationship among tourists, attraction tags, and the attractions themselves is utilized to develop the tourist preference model. The second step involves calculating the similarity between tourist attractions, which measures how alike two different attractions are. Finally, a personalized recommendation set for tourist attractions is generated based on the tourist's interests and the calculated similarities. The system computes the degree of preference a tourist has for each attraction and formulates the recommendation result set accordingly.

In this study [21], addressed the limitations of generalized packages offered by travel agencies, which restrict tourists' ability to choose their preferred options. They incorporated content-based filtering in the review module of their RS. The proposed system is built on three algorithms. The Euclidean Distance algorithm calculates the distance between two locations. the k-nearest neighbors (KNN) algorithm identifies the nearest attractions and locations, using the Euclidean distance formula to determine which places are within a specified range. Lastly, the apriori algorithm is employed for classification. The system recommends new destinations to tourists, suggests nearby attractions and frequently visited places, provides reviews for these locations, offers a hotel booking interface, and supplies route information.

In this study [22], developed a user-based recommender system for tourist attractions. This system functions as an online application capable of generating a personalized list of preferred attractions for tourists. The process involves three main steps: representing user (tourist) information, analyzing and modeling the visiting history of attractions, and generating a list of similar users (tourists). The similarity between tourists is calculated based on their visiting history data using a collaborative filtering algorithm. The system then generates recommendations for the top-N attractions tailored to the tourist, informed by the visiting history of their neighbors. By utilizing users' basic information and past travel histories, the system identifies a list of neighbor users recorded in the user database. When users log into the system, they receive attraction recommendations based on the travel experiences of their neighbors.

In this study [23], created a service with web and mobile interfaces designed to help users discover previously unknown tourist POI. They integrated a collaborative filtering-based recommendation engine that generates suggestions based on locations the user has rated, as well as ratings from other users. The service allows users to visualize attractions near their current geographic location, facilitating the effortless discovery of new places. The developed solution features an intuitive interface compatible with widely used iOS platforms. Usability and load tests conducted on the application yielded satisfactory results. This application aims to enhance tourism by promoting a variety of attractions, including lesser-known sites, in the vicinity of the user's location.

Meehan *et al*. [24] addressed issues related to traditional tour guide applications, which primarily focus on location while neglecting other contextual factors. This limitation has resulted in inappropriate suggestions due to insufficient content filtering and has led to tourists experiencing information overload. The study employs a context-aware hybrid recommendation technique, utilizing five main types of contextual data: location, time, weather, social media sentiment, and personalization. By incorporating these diverse contextual factors, the application yields more accurate results for users. After examining various recommendation techniques, the study concluded that the hybrid recommendation approach is the most suitable solution for the identified research problem.

The study [25] introduced collaborative filtering to a new domain, visit personalization, using a simple yet efficient and scalable approach. It gathers tourist preferences in a non-intrusive manner by leveraging a user's public tagging records and evaluates tourist personalization techniques based on actual tourist experiences. The attractions a person chooses to visit are influenced not only by the popularity of landmarks but also by individual preferences. The authors analyze the records of visited landmarks from online user data to create a user-user similarity matrix. When a user expresses interest in a new destination, the system generates a list of potentially appealing attractions based on the experiences of similar users who have already visited that location. A key finding of the study is that accurate photo tagging enhances the experience not only for other users but also for the target user, who benefits from personalized tourist recommendations. Table 1 provides a summary of related studies focusing on tourism RS.

Table 1. A summary of recommendation system studies

Study	Proposed approach	Recommendation techniques	Dataset	Evaluation metric
[4]	Personalized recommendation system based on the user's preferences, diversity and distance.	Collaborative filtering	TripAdvisor and Naver	User ratings and metadata of POIs
[12]	Evolution of context-aware personalized travel recommender system	Content-based filtering Collaborative filtering	_	Performance metrics Prediction quality metrics Ranking metrics
[13]	Recommendation system using Semantic clustering and sentiment analysis	User-based filtering	A dataset gathered from TripAdvisor platform	Similarity metric Symmetric matrix
[14]	Framework design for personalized recommendations of online tourism	Collaborative filtering Content-based	_	Analysis of various accuracy metrics
[15]	Incorporation of user reviews into tourist RS.	Content-based	Europe hotels	_
[16]	Contextual information for RS	Hybrid filtering	Foursquare	LCS
[18]	Genetic algorithm and LSTM network to forecast the popularity and ratings of tourism services consider user influence within social networks and time series data.	Collaborative filtering	Real-world dataset	Analysis of various accuracy metrics
[19]	Hybrid recommendation based on new user cold-start and new package cold-start algorithms	Hybrid filtering and demographic filtering (DF)	Tripadvisor website for Thailand POIs	User similarity matrix
[20]	Personalized recommendation based on tags algorithm and few algorithm	Tag and collaborative filtering	100 Pieces of user data of a tourism website	User similarity, tourist-attractions tag, tourists' interest preference
[21]	Recommendation system using data mining techniques	Content based filtering	_	Recall metric
[22]	Online application to generate personalized list of attractions for the Tourist	User-based collaborative filtering	_	User similarity
[23]	Tourist guide integrated with a recommender system that facilitates interaction among users.	Collaborative filtering	Movie Lens	Mean absolute error
[24]	Context-aware recommendation system	Hybrid filtering	Londonderry city's attractions filtering	_
[25]	Personalized recommendation system developed through the analysis of social media data.	Collaborative filtering	Flickr users	User similarity matrix

From our review, we can draw the following conclusions:

- Most studies have concentrated on collaborative filtering as the primary recommendation technique. This approach is driven by the notion that users tend to receive the best recommendations from others who share similar preferences. Furthermore, collaborative filtering enables the model to assist users in discovering new interests without requiring extensive domain knowledge.
- Certain studies, including [16], [19], and [24], have explored hybrid filtering techniques. They incorporated contextual information, such as geographical and temporal data, to enhance and assess the quality of recommendations.
- Only the study [20] focused on integrating collaborative filtering with tagging. We believe that attraction tags allow tourists to express their opinions about various attractions. Thus, these tags serve as a crucial link between tourists and tourist sites, providing valuable data to address their interests. Each tourist can tag multiple attractions simultaneously, and the collection of tags assigned by each tourist can reveal their preferences and hobbies.

4. PROPOSED RECOMMENDATION SYSTEM

This study introduces a personalized tourism RS specifically designed for the Qassim region in the Kingdom of Saudi Arabia. The proposed system utilizes a tag-based approach combined with collaborative filtering techniques to recommend tourism options in various cities within Qassim. It encompasses all places and POIs in the region, providing tailored recommendations based on users' stored preferences. This system enables tourists to plan their trips more effectively while minimizing the time and effort required to find locations that align with their interests.

4.1. System methodology

The proposed system comprises two primary components. Figure 2 depicts the overall methodology of this study. Below, we provide a detailed description of the various phases of the methodology.

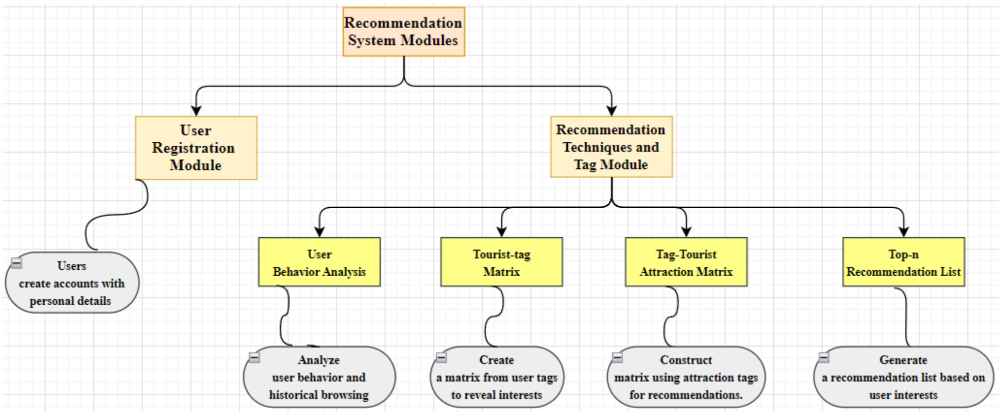

Figure 2. Proposed modules for the recommendation system

4.1.1. User registration module

In this module, users are required to create an account within the system. During the registration process, they must provide personal details including their name, date of birth, address, phone number, email address, and password. To ensure user privacy, this information will be stored in an encrypted database. Once the registration is complete, users can log into the system. User registration module includes the following steps:

- User login: when a user attempts to log in, they are prompted to enter their username/email and password. If the credentials match, the user gains access to the system; if not, an error message is displayed, prompting them to retry.
- Data retrieval: upon successful login, the system retrieves the user's profile, including their stored preferences, historical data, and any tags associated with their past interactions. This data is essential for generating personalized recommendations.
- Security measures: all personal information, including passwords, is stored in an encrypted format to protect user privacy.

This module is designed to provide users with a secure and user-friendly interface for accessing and managing their personal information, ensuring that their data is both accessible and protected. By facilitating easy login and data management, the module plays a vital role in enhancing user experience and engagement with the tourism RS.

4.1.2. Recommendation techniques and tag module

In this module, tourists utilize attraction tags to express their opinions about various tourist sites. These tags serve as a connection between tourists and attractions, providing valuable data to cater to tourists' interests. Each tourist can tag multiple attractions simultaneously. The collection of tags assigned by each tourist can reveal their preferences and hobbies. In this module, first, the user's behavior and historical browsing records within the system will be analyzed. Based on this data, a tourist-tag matrix will be created to develop an interest model for tourists regarding various attractions, achieved by constructing a tag-tourist attraction matrix. This matrix will be built using the set of attraction tags. To compute the similarity between tourist attractions, it is noted that different tourists may assign various tags to the same attractions. The greater the number of tags assigned to an attraction, the more characteristics of that attraction can be represented, which will be calculated using a similarity metric. Lastly, a top-n recommendation list will be generated based on the tourist interest model and the similarity among the attractions.

Recommendation techniques and tag module includes the following steps:

- User behavior analysis: the system analyzes user behavior and historical browsing records to understand how tourists interact with different attractions. This analysis helps identify patterns in tagging behavior, revealing which types of attractions resonate most with particular user groups.
- Tourist-tag matrix creation: based on the collected tags, a tourist-tag matrix is created to model the interests of each tourist in relation to various attractions. This matrix reflects how frequently and which tags are assigned to each attraction by different tourists.
- Tag-tourist attraction matrix: the system constructs a tag-tourist attraction matrix that maps tags to specific

attractions. This matrix provides insight into the collective interests of tourists, enabling more tailored recommendations. To enhance the recommendation process, the system calculates the similarity between attractions based on the tags assigned by different tourists. Attractions with a greater number of tags are better represented in terms of their characteristics, allowing for more accurate comparisons.

- Top-n recommendation list: finally, a top-n recommendation list is generated using the tourist interest model and the calculated similarities between attractions. This list provides tourists with personalized suggestions that align with their interests and preferences, enhancing their experience and increasing the likelihood of exploration.

This module leverages user-generated content to create a dynamic and responsive RS that caters to the diverse interests of tourists.

4.2. System design

A data flow diagram (DFD) has been utilized to outline the architecture of the proposed RS. Figure 3 illustrates the system design, which will be explained in detail below.

The components of a DFD include the following:

– External entities: these represent users, such as tourists who interact with the system. They provide input (such as personal information and attraction tags) and receive recommendations.
– Processes: the main functions of the system, such as user registration, tag input, attraction tagging, and recommendation generation, are depicted as processes. Each process transforms the input data into meaningful output.
– Data stores: these indicate where the system's data is stored, including user profiles, attraction tags, and the Tourist-tag matrix. Data stores are crucial for maintaining historical records and user preferences.
– Data flows: arrows illustrate how data moves between external entities, processes, and data stores. This visual representation helps clarify how information is gathered, processed, and used to generate recommendations.

By analyzing the DFD, we can gain insights into the interactions within the RS and understand how user input leads to tailored suggestions for tourist attractions.

The user registration and login module functions as a crucial component of the proposed personalized tourism RS. Here's a detailed breakdown of its operations: when a user attempts to log in, the user registration and login module retrieves their credentials, including their username/email and password, from the database created during their initial registration. Additionally, users have the option to update their information, such as passwords, email addresses, and personal details.

The system also monitors the user's previous browsing activity and behavior to support the tag and recommendation modules. This data will be used to develop the tourist-tag framework, which analyzes the level of interest that tourists display in travel-related attractions. A tag-tourist attraction matrix will be constructed using the set of attraction tags. The similarity of tourist attractions can be assessed, as different visitors may assign various tags to the same attraction. The more tags assigned to an attraction, the more features can be represented, and similarity metrics will facilitate this analysis. Finally, a top-n recommendation list will be generated based on the tourist interest model and the calculated similarities among the attractions.

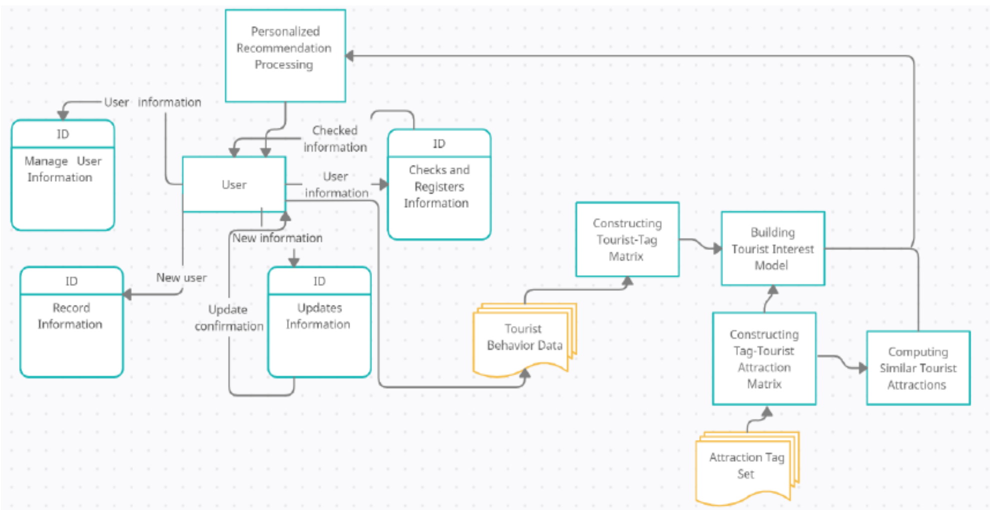

Figure 3. System design using data flow diagram

5. SYSTEM EVALUATION

5.1. Data collection

To evaluate the performance of the proposed RS, it is crucial to utilize a real dataset. Therefore, we developed a dataset specifically for the Qassim region, which includes various files containing information about users and different attractions. This dataset encompasses details on: i) user information, ii) attractions, iii) chalets, iv) heritage landmarks, v) parks, vi) restaurants, vii) public places, viii) malls, and ix) hotels

The data was gathered from multiple sources, including Google Maps and Instagram, ensuring a comprehensive representation of the region's offerings. This rich dataset will enable us to analyze the effectiveness of the RS and its ability to cater to user preferences and interests. Some samples from the gathered dataset are illustrated in the following formats:

- Table 2 provides a description of the attraction file, showcasing various attributes and details relevant to the attractions in the Qassim region.
- Figure 4 displays a sample of the chalets, highlighting key information such as location, amenities, and user ratings.
- Figure 5 presents a sample of heritage landmarks, offering insights into their historical significance, features, and visitor information.

These representations will help in understanding the structure and content of the dataset used for developing the RS.

Table 2. Attraction dataset's attributes and their descriptions

Attribute name	Attribute description
Category	Classification of attractions for specific categories
Title	Attractions' titles
URL	Link to all attractions in the Google Map application
Item ID	Item ID assign special id for each item

Chalet name	Region	Rating	Instagram accounts
Garlands Chalet	Unayzah	3.6	
Laberla Chalet	Buraydah	4.2	https://instagram.com/laperlachalet?utm_medium=copy_link
Nostalgia Chalet	Buraydah	3.9	
Dream Chalet	Buraydah	4.7	https://instagram.com/dreem_shalih?utm_medium=copy_link
Le Porto Chalet	Unayzah	4.6	https://instagram.com/chalet_le_porto?utm_medium=copy_link
Avenues Chalet	Unayzah	3.5	https://instagram.com/avino.chalet?utm_medium=copy_link
The White Dream Chalet	Unayzah	4.5	https://instagram.com/_thewhitedream_?utm_medium=copy_link
Tamara Chalet	Unayzah	4.6	https://instagram.com/tamara_unaizah?utm_medium=copy_link
Family Day Chalet	Unayzah	4.3	https://instagram.com/family_day__?utm_medium=copy_link
Diamond Chalet	Unayzah	4.2	https://instagram.com/cha_almas?utm_medium=copy_link
Papel Chalet	Riyadh Al-I	4.4	https://instagram.com/babylon_chalet?utm_medium=copy_link
Riva Chalet	Riyadh Al-I	4.8	https://instagram.com/revaa2018?utm_medium=copy_link
Elite Chalet	Al-Rass	4	
Boulevard Chalet	Al-Rass	4.1	
Lavender Chalet	Al-Rass	4.7	https://instagram.com/shalyh_lafnder?utm_medium=copy_link

Figure 4. Sample of the Qassim region chalets

Heritage landmarks	Region	Rating
Al-Oqilat Museum	Buraydah	4.5
Al-Musawkaf Market	Unayzah	4.2
Al-Shinanah Tower	Al-Rass	4.1
Maqsorat Al-Swailem	Al-Bukayiryah	4.2
Al-Debikhi Museum	Buraydah	3.9
Jedaya heritage castle	Al-Rass	4.4

Figure 5. Sample of Qassim region heritage landmarks

5.2. Data preprocessing

To ensure the accuracy and quality of the data, data preprocessing is essential. This section outlines the various steps involved in the data preprocessing phase. The automatic cleaning processes are carried out using Python code, and these steps can be summarized as follows:

- Checking null values and drop Instagram account: in this stage, we check null values from the dataset and we drop Instagram accounts from the records as shown in Figure 6.
- Removing records with missing values: after checking null values, we remove the records where the region is missing in the data frame as shown in Figure 7.

– Calculating the average rating: in this stage, we used the mean function to determine the mean of each place by taking the ratings of the first five users for each place, repeating the process five times with a different rating for each place. The average rating for Heritage landmarks is shown in Figure 8.

These preprocessing steps are crucial for preparing the data for subsequent analysis and modeling, ultimately leading to more accurate and reliable recommendations within the proposed system.

```
In [5]: #check null data
        data.isnull().sum()

Out[5]: Chalet name          0
        Region               0
        Rating               0
        Instagram accounts   4
        dtype: int64

In [6]: data.drop('Instagram accounts', axis = 1, inplace = True)

In [7]: data
Out[7]:
```

	Chalet name	Region	Rating
0	Garlands Chalet	Unayzah	3.6
1	Laberla Chalet	Buraydah	4.2
2	Nostalgia Chalet	Buraydah	3.9

Figure 6. Checking null values and drop Instagram accounts

```
In [8]: #check null data
        data.isnull().sum()

Out[8]: Chalet name    0
        Region         0
        Rating         0
        dtype: int64

In [10]: # Dropping the records with Region missing in data dataframe.
         data = data[~data.Region.isnull()].copy()

In [11]: data
Out[11]:
```

	Chalet name	Region	Rating
0	Garlands Chalet	Unayzah	3.6
1	Laberla Chalet	Buraydah	4.2
2	Nostalgia Chalet	Buraydah	3.9
3	Dream Chalet	Buraydah	4.7
4	Le Porto Chalet	Unayzah	4.6

Figure 7. Removing records with missing values

```
In [41]: #Final with average repeated records
         data.groupby('Heritage landmarks').mean()

Out[41]:
```

Heritage landmarks	Rating
Al-Debikhi Museum	4.14
Al-Musawkaf Market	3.54
Al-Oqilat Museum	4.04
Al-Shinanah Tower	4.16
Jedaya heritage castle	4.06
Maqsorat Al-Swailem	3.74

Figure 8. Average rating for heritage landmarks

5.3. System prototype

We have created a personalized tourist RS for the Qassim region using Python and Android Studio tools. The system prompts users to select from a list of their interests, enabling it to provide tailored recommendations based on their preferences. Below, we present some screenshots of the proposed system. Figure 9 shows the booking screen. The locations in the popular section will be displayed according to the tags selected by the user in the interest's screen.

Figure 9. Booking screen

When a user selects a location from the RS, detailed information about the selected location is displayed. This includes an overview of the location, its key features, and any additional details such as images, user reviews, and nearby attractions. If the user decides to proceed, they can initiate a reservation by clicking the “Book” button. This action directs the user to a new screen where they are prompted to choose their preferred date and time for the visit. Once the date and time are selected, the user can review the information they have entered. To finalize the reservation, they simply click the “Confirm” button, completing the booking process.

Additionally, users have the option to create customized trips through the trip creation feature, as illustrated in Figure 10. On the trip creation screen, the user is required to provide specific details about the trip. They start by giving the trip a unique name that reflects its purpose or theme. Next, they select the destination from the available options and specify the maximum number of participants who can join the trip. The user also defines the trip’s start and end times, ensuring a clear schedule for the planned activities. This comprehensive trip creation process allows users to organize their travel plans effectively and share the experience with others if desired.

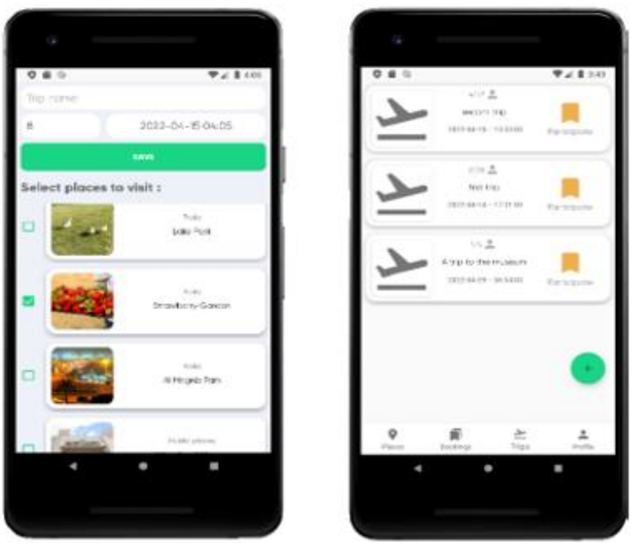

Figure 10. Trip creation screen

The trip screen displays the possible trips that the user may be interested in, along with a brief description of the location, the departure time, and the maximum number of participants. By selecting the Plus Sign in the upper right corner of the screen, users can also create a trip. These screenshots demonstrate the user interface and functionality of the system, showcasing its user-friendly design and personalized recommendation capabilities.

5.4. Discussion

Based on the system evaluation, the following findings can be derived:

– We developed a customized RS specifically for the Qassim region of Saudi Arabia. This system allows users to select their interests from a list, then provides tailored recommendations, helping tourists design a suitable itinerary by efficiently collecting, assessing, and displaying relevant information.
– The proposed system saves users time and effort by offering personalized travel recommendations, enabling them to tailor their schedules according to personal preferences while avoiding destinations that may not align with their interests.
– The RS methodology is based on collaborative filtering and attraction tags. Attraction tags enable users to share their opinions about various sites, effectively creating a connection between tourists and travel destinations. These tags provide rich insights into visitor interests, allowing the system to deliver personalized recommendations. Additionally, travelers can tag multiple destinations, revealing details about their hobbies and preferences.
– We have compiled a real dataset specifically for the Qassim region. This dataset includes multiple files covering user information, attractions, chalets, heritage landmarks, parks, restaurants, public places, malls, and hotels. To ensure data accuracy, we performed data preprocessing, with each stage detailed in this study.

Our findings indicate that user-selected interests significantly correlate with the relevance of recommended travel destinations. The proposed method in this study yielded a notably higher proportion of itinerary satisfaction, as users were able to tailor their experiences to match specific preferences. Additionally, the use of attraction tags proved effective in aligning recommendations with user interests, as these tags served as valuable indicators of visitor preferences, resulting in more targeted and personalized recommendations. Data preprocessing on the newly created dataset for the Qassim region further enhanced accuracy, ensuring that recommendations were grounded in well-organized and relevant data. This study explored a comprehensive personalized RS for the Qassim region with a focus on tailoring itineraries to user preferences. However, further in-depth studies may be necessary to confirm its effectiveness across diverse tourist demographics, especially regarding the adaptability of the recommendation algorithm to evolving user interests over time. Additionally, while attraction tags proved useful, a more extensive dataset covering a broader range of activities and seasonal trends may be required to ensure recommendations remain relevant and responsive to fluctuating tourism patterns.

6. CONCLUSION

Tourism presents a significant opportunity for the application of recommender systems, as tourists often face an overwhelming array of attractions and must invest considerable time and effort in selecting the most suitable options. To address this challenge, we proposed a tag-based RS utilizing collaborative filtering, specifically tailored for tourists in the Qassim region of the Kingdom of Saudi Arabia. The primary objectives of this system are to enhance tourism in the area and provide highly accurate recommendations for travelers. Initially, our study focused on examining the tourism RS and analyzing popular literature-based recommendation techniques. We conducted a comprehensive review of existing research and proposed methods to determine the most effective technology for our system. Additionally, we discussed the strengths and weaknesses of each approach. Following this analysis, we outlined the design of the system, detailed the dataset we created, and described the various stages of development. The proposed system allows users to select from a list of their interests, subsequently displaying relevant recommendations tailored to their preferences. In conclusion, recent observations suggest that tourists in the Qassim region face significant challenges in selecting suitable attractions amidst an overwhelming array of options. Our findings provide conclusive evidence that the implementation of a tag-based RS can effectively enhance the travel experience by delivering personalized and relevant suggestions tailored to individual preferences. This improvement is not merely a result of increased data availability but rather reflects the system’s ability to align recommendations with user interests, ultimately leading to higher satisfaction and engagement in the region’s tourism offerings.

Looking ahead, we plan to broaden our dataset to include a wider variety of tourist destinations and to introduce a user-comment section for evaluation and rating purposes. Moreover, we intend to create a hybrid

RS that integrates content-based filtering, collaborative filtering, and demographic information. This strategy aims to improve recommendation outcomes by providing users with a wider selection of options and predicting ratings for new users, allowing for the identification of similar items through neighborhood-based techniques. Our findings indicate that personalized RS designed around user interests significantly improve traveler satisfaction when compared to generic models. Future research could investigate the incorporation of machine learning techniques to enhance these recommendations, particularly in terms of dynamically adjusting to evolving user preferences. Additionally, examining the effects of cultural and seasonal influences on tourist preferences could offer valuable insights for developing more robust and adaptable RS that cater to a diverse range of visitor demographics.

ACKNOWLEDGEMENTS

The author(s) gratefully acknowledge Qassim University, represented by the Deanship of "Scientific Research, on the financial support for this research under the number (COC-2022-1-3-J-31639) during the academic year 1444 AH / 2022 AD".

REFERENCES

[1] S. Nabil, M. Y. Chkouri, and J. El Bouhdidi, "Demographic information combined with collaborative filtering for an efficient recommendation system," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 5, pp. 5916–5925, 2024, doi: 10.11591/ijece.v14i5.pp5916-5925.

[2] S. Gupta and M. Dave, "An overview of recommendation system: methods and techniques," *Advances in Computing and Intelligent Systems: Proceedings of ICACM*, 2020, pp. 231–237, doi: 10.1007/978-981-15-0222-4_20.

[3] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modelling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002, doi: 10.1023/A:1021240730564.

[4] J. Lee, J. A. Shin, D. K. Chae, and S. C. Lee, "Personalized tour recommendation via analyzing user tastes for travel distance, diversity and popularity," *Electronics (Switzerland)*, vol. 11, no. 7, p. 1120, Apr. 2022, doi: 10.3390/electronics11071120.

[5] P. Sirohi *et al.*, "Energy-efficient cloud service selection and recommendation based on QoS for sustainable smart cities," *Applied Sciences (Switzerland)*, vol. 11, no. 20, p. 9394, Oct. 2021, doi: 10.3390/app11209394.

[6] S. Li and J. An, "Things of interest recommendation with multidimensional context embedding in the internet of things," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 165–174, 2023, doi: 10.14569/IJACSA.2023.0140517.

[7] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proceedings of the ACM International Conference on Digital Libraries*, Jun. 2000, pp. 195–204. doi: 10.1145/336597.336662.

[8] H. Pandey and V. K Singh, "A fuzzy logic based recommender system for E-learning system with multi-agent framework," *International Journal of Computer Applications*, vol. 122, no. 17, pp. 18–21, Jul. 2015, doi: 10.5120/21793-5140.

[9] H. Jafarkarimi, A. Tze Hiang Sim, and R. Saadatdoost, "A naive recommendation model for large databases," *International Journal of Information and Education Technology*, vol. 2, no. 3, pp. 216–219, 2012, doi: 10.7763/IJIET.2012.V2.113.

[10] Y. Wang, Z. Qin, J. Tang, and W. Zhang, "Optimization of digital recommendation service system for tourist attractions based on personalized recommendation algorithm," *Journal of Function Spaces*, vol. 2022, pp. 1–9, Sep. 2022, doi: 10.1155/2022/1191419.

[11] Z. Jia, Y. Yang, W. Gao, and X. Chen, "User-based collaborative filtering for tourist attraction recommendations," in *Proceedings - 2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*, Feb. 2015, pp. 22–25. doi: 10.1109/CICT.2015.20.

[12] S. Renjith, A. Sreekumar, and M. Jathavedan, "An extensive study on the evolution of context-aware personalized travel recommender systems," *Information Processing and Management*, vol. 57, no. 1, p. 102078, Jan. 2020, doi: 10.1016/j.ipm.2019.102078.

[13] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Systems with Applications*, vol. 167, p. 114324, Apr. 2021, doi: 10.1016/j.eswa.2020.114324.

[14] X. Wang, "Personalized recommendation framework design for online tourism: know you better than yourself," *Industrial Management and Data Systems*, vol. 120, no. 11, pp. 2067–2079, Oct. 2020, doi: 10.1108/IMDS-05-2020-0278.

[15] K. Kesorn, W. Juraphanthong, and A. Salaiwarakul, "Personalized attraction recommendation system for tourists through check-in data," *IEEE Access*, vol. 5, pp. 26703–26721, 2017, doi: 10.1109/ACCESS.2017.2778293.

[16] P. Sánchez, "Exploiting contextual information for recommender systems oriented to tourism," in *RecSys 2019 - 13th ACM Conference on Recommender Systems*, Sep. 2019, pp. 601–605. doi: 10.1145/3298689.3347062.

[17] R. Hassannia, A. V. Barenji, Z. Li, and H. Alipour, "Web-based recommendation system for smart tourism: multiagent technology," *Sustainability (Switzerland)*, vol. 11, no. 2, p. 323, Jan. 2019, doi: 10.3390/su11020323.

[18] J. Ye, Q. Xiong, Q. Li, M. Gao, and R. Xu, "Tourism service recommendation based on user influence in social networks and time series," in *Proceedings - 21st IEEE International Conference on High Performance Computing and Communications, 17th IEEE International Conference on Smart City and 5th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2019*, Aug. 2019, pp. 1445–1451. doi: 10.1109/HPCC/SmartCity/DSS.2019.00200.

[19] M. L. Bai, R. Pamula, and P. K. Jain, "Tourist recommender system using hybrid filtering," in *2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019*, Nov. 2019, pp. 746–749. doi: 10.1109/ISCON47742.2019.9036308.

[20] Y. Cui, C. Huang, and Y. Wang, "Research on personalized tourist attraction recommendation based on tag and collaborative filtering," in *Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019*, Jun. 2019, pp. 4362–4366. doi: 10.1109/CCDC.2019.8832961.

[21] S. Kokate, A. Gaikwad, P. Patil, M. Gutte, and K. Shinde, "Traveler's recommendation system using data mining techniques," in *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, Aug. 2018, pp. 1–5. doi: 10.1109/ICCUBEA.2018.8697862.

[22] Z. Jia, Y. Yang, W. Gao, and X. Chen, "User-based collaborative filtering for tourist attraction recommendations," in *Proceedings - 2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*, Feb. 2015, pp. 22–25. doi: 10.1109/CICT.2015.20.
[23] A. Umanets, A. Ferreira, and N. Leite, "GuideMe – a tourist guide with a recommender system and social interaction," *Procedia Technology*, vol. 17, pp. 407–414, 2014, doi: 10.1016/j.protcy.2014.10.248.
[24] K. Meehan, T. Lunney, K. Curran, and A. McCaughey, "Context-aware intelligent recommendation system for tourism," in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2013*, Mar. 2013, pp. 328–331. doi: 10.1109/PerComW.2013.6529508.
[25] A. Popescu and G. Grefenstette, "Mining social media to create personalized recommendations for tourist visits," in *ACM International Conference Proceeding Series*, 2011, pp. 1–6. doi: 10.1145/1999320.1999357.

BIOGRAPHIES OF AUTHORS

Afef Selmi received the master's and Ph.D. degrees in computer science from the Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia, in 2011 and 2017, respectively. From 2011 to 2017, she worked as a research assistant with the University of Sousse, Tunisia. Since 2018, she has been an assistant professor with the College of Computer, Qassim University, Saudi Arabia. She has attended and participated in various international conferences. Her research interests include recommendation systems, data mining, and machine learning. She received the Publication Award from Qassim University, in 2020. She can be contacted at email: a.selmi@qu.edu.sa.

Maryah Alawadh received the B.Sc. degree in information technology, College of Computer, Qassim University, Saudi Arabia in 2022. Her research interests include data science, recommendation system, and natural language processing. She can be contacted at email: 372206731@qu.edu.sa.

Raghad Alotaibi received the B.Sc. degree in information technology, College of Computer, Qassim University, Saudi Arabia in 2022. Her research interests include artificial intelligence, machine learning and data analysis. She can be contacted at email: 381211830@qu.edu.sa.

Shrefah Alharbi received the B.Sc. degree in information technology, College of Computer, Qassim University, Saudi Arabia in 2022. Her research interests include artificial intelligence, machine learning. She can be contacted at email: 381225292@qu.edu.sa.