Recommendation System on Travel Destination based on Geotagged Data

Clarice Wong Sheau Harn, Mafas Raheem
School of Computing,
Asia Pacific University of Technology and Innovation, Technology Park, Kuala Lumpur, 57000, Malaysia

***Abstract*—Tourism research has benefitted from the worldwide spread and development of social networking services. People nowadays are more likely to rely on internet resources to plan their vacations. Thus, travel recommendation systems are designed to sift through the mammoth amount of data and identify the ideal travel destinations for the users. Moreover, it is shown that the increasing availability and popularity of geotagged data significantly impacts the destination decision. However, most current research concentrates on reviews and textual information to develop the recommendation model. Therefore, the proposed travel recommendation model examines the collective behaviour and connections between users based on geotagged data to provide personalized suggestions for individuals. The model was developed using the user-based collaborative filtering technique. The matrix factorization model was selected as the collaborative filtering technique to compute user similarities due to its adaptability in dealing with sparse rating matrices. The recommendation model generates prediction values to recommend the most appropriate locations. Finally, the model performance of the proposed model was assessed against the popularity and random models using the test design established using Mean Average Precision (MAP), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The findings indicated that the proposed matrix factorization model has an average MAP of 0.83, with RMSE and MAE values being 1.36 and 1.24, respectively. The proposed model got significantly higher MAP values and the lowest RMSE and MAE values compared to the two baseline models. The comparison shows that the proposed model is effective in providing personalized suggestions to users based on their past visits.**

***Keywords*—Geotagged data; travel recommendation system; travel recommender; collaborative filtering; matrix factorization**

I. Introduction

A. Background

The World Travel and Tourism Council reported that the tourism industry contributed around 11% to the global economy in 2019 before the pandemic outbreak [1]. This illustrates that the tourism industry has been one of the most influential and profitable industries and has been a major contributor to the global economy. It is also one of the most promising sectors. The travel industry has evolved dramatically over time, and the introduction of big data analytics has profoundly impacted people's travel. Travelers used to rely on newspapers, magazines, and radio to get to know about places and plan their trips with help of the travel agencies. However, in modern society, travelers have numerous options to plan their trips.

The rise of big data and the evolution of technology have significantly impacted people's travel [2], [3]. Today, many people book their trips online using platforms such as TripAdvisor and Expedia. The increasing amount of data collected and accessible by travel providers has facilitated the creation of sophisticated analytics and forecasting algorithms. The rise of social media has also greatly impacted how people communicate allowing users to exchange content, such as pictures and videos, and has greatly aided in human interaction. The report by Wyman [4] further illustrates that travelers have expanded their social media usage by 44% since the pandemic, and 92% of users find useful information online about places to visit.

The impact of social media on travel destinations was investigated in several studies in the past. Various social media platforms, blogs and online communities are becoming more prevalent in the travel industry as they allow users to connect and share their experiences [5-7]. Social media platforms such as Twitter, Flickr, and Facebook have enabled travelers to share information and express their travel experiences online, which has helped boost the reputation of a city as a desirable travel destination. The study also noted that Generation Z and Millennials are more likely to utilize social media to plan their vacations because they value online experiences more than commercials. The study [8] elaborated that having a strong online presence is very important for a destination to gain a positive reputation, and social media significantly impacts how customers choose travel destinations.

The travel industry is one of the most data-driven industries in the world due to the exponentially increasing amount of data. It is often difficult for individuals to select the ideal holiday destination due to a lack of understanding of the various attractions and the complexity of the planning process. Consequently, several research works were conducted on travel recommendations that consider the different elements to deliver personalized recommendations based on the preferences and behaviours of users [9-11].

A geotagging service is a type of geographic identification service that may be used to identify the location of a media file or social media post. The data typically includes a latitude and longitude coordinate that may be used to locate the captured place on a map, as well as the date and time the picture or post was filmed [12]. According to the study, the growth of location-based social networks has enabled individuals to construct their social networks based on interpersonal contacts. Studies explained that the expansion in publicly available geotagged social media data may be attributed to the adoption of social media platforms such as Facebook, Instagram, and Flickr [13-15]. In addition, geotagged data is rich in information about the users' interests and may be utilized to discover new regions of interest. The statement is reinforced and demonstrated that the analysis of geotagged data may assist the government in promoting tourist places [16]. Therefore, using geotagged social media data to identify tourist hotspots is advantageous for developing a trip recommendation model.

A major portion of establishing successful travel recommendation systems has focused on examining the reviews from Google Maps and TripAdvisor. Although the models are accurate, oriented towards a few famous landmarks and do not utilize geotagged data to make personalized recommendations depending on the user's preferences.

B. Problem Statement

Numerous studies indicate that the prevalence of social media networks that supply geotagged data is increasing the influence of this data type on users' destination selections [5], [7], [17]. Due to the rising popularity of geotagged social media as a source of travel destinations, the existing travel recommendation system lacks the efficiency to fully leverage the information gathered from geotagged data to construct a travel recommendation model that can study users' preferences.

C. Aim

The research aims to determine the role of geotagged data in shaping the selection of attractions using clustering techniques and develop a travel recommendation model as well as compare it to the models with different approaches.

D. Significance and Scope

Extensive studies have attempted to develop a personalized travel recommendation model to assist people in filtering data to attractions that match their preferences [10], [18], [19-20]. The proposed travel recommendation system allows travelers to discover possible destinations based on their interests. Additionally, the tourist sector may utilize the potential information gathered from the research to establish successful marketing plans in the tourism sector and increase operational efficiency. The dataset for this research was acquired from Kaggle which contains 20,000 geotagged data points in London between 2014 and 2019 [21]. London is recognized for its various attractions, including stunning architecture and historical buildings. Given that the coronavirus pandemic in 2020 may impede travel mobility, the final year of 2019 in the dataset is ideal [22]. Given that the data obtained for this research was collected at random and across time, the recommendation model will focus on location recommendations rather than route suggestions.

II. Related Works

A brief history of the recommendation system describes various forms of recommendation systems including the current state-of-the-art techniques for the travel recommendation system. Past studies conducted using geotagged data were reviewed and summarized at the end of this section.

A. Recommendation System

During the 1990s, research in the field of recommendation models started focusing on developing systems that can predict product ratings [23-24]. The recommendation model suggests the best suitable goods and services to its consumers through the information gathered [25]. The most notable examples include Amazon's personalized shopping system, YouTube's suggestions for videos relevant to the viewers' interests, and Facebook's system allows users to interact with more people. The rise of information technology made users and organizations more dependent on recommendation systems to sift through the vast amount of data collected in this century.

B. Types of Recommendation Systems

The recommendation systems can be categorized into four different categories such as content-based, collaborative-based, hybrid-based, and knowledge-based [25-30]. The content-based method connects user traits with items most likely to satisfy their demands. The collaborative filtering technique, on the other hand, believes that individuals with similar interests and historical behaviours would act similarly in the future. The disadvantage of implementing a content-based system is that it heavily relies on the knowledge base to provide recommendations while implementing collaborative filtering can be very challenging when the users are relatively new to the platform. A knowledge-based system can be very useful in helping users find the best products and services that are seldom acquired, such as luxury goods and real estate. As the calculation is based on the similarity between the item descriptions and the user's needs, it is essential to outline the knowledge base needed to generate recommendations [31]. However, the process can be very time-consuming and costly. Besides, a hybrid recommendation system uses the best elements from various techniques to overcome these shortcomings. The research [29] proposed the hybrid system integrated the characteristics of content-based and collaborative filtering to have the right predictions.

To date, numerous pieces of literature have studied the recommendation system in a broad range of industries. For instance, [30] proposed a matrix factorization-based recommender system that can recommend books based on the similarities and ratings of users. In contrast, a movie recommendation system was developed by applying the content filtering technique to exploit the movie's genre characteristics and requested the users to answer a few questions while building up their profiles to enrich the users' data [32]. Moreover, [33] used the content-based approach to suggest music to users based on the musical features utilizing the convolutional recurrent neural network (CRNN) method. As online learning sites grew in popularity, a knowledge-based recommendation system was proposed, which serves as an agent to assist users in recommending appropriate courses and materials based on their preferences and requirements [34].

C. Algorithms used in Travel Recommendation System

Multiple studies on recommender systems in the tourism industry were undertaken over the years to assist individuals with their trip decisions. The authors in [35] analyzed the essential information and interests of the users on social media platforms to suggest travel-related activities in Tunisia. The proposed system created user profiles using the content-based filtering algorithm based on the content they posted on social media sites such as Facebook, TripAdvisor, and Twitter. Furthermore, [11] conducted a study that analyzed the opinions of Korean undergraduates on various destinations. The study used a collaborative method to analyze the similarities and differences between the users. It also considered the various restrictions and demands to provide personalized suggestions. However, since the data for the study was only collected from a single university, the results might be biased.

The study [10] collected users' reviews from TripAdvisor and developed a collaborative system to deliver suggestions to their users. The researchers used a combination of text processing and semantic clustering to analyze the data and extract their preferences for recommendations. However, the research only acquired 100 reviews from the site on specific locations, which may have led to skewed findings and the neglect of other less-visited attractions. Likewise, [20] sought to discover the ideal tour route for international visitors in South Korea by examining TripAdvisor ratings. Text mining and network analysis were used to perform a comprehensive analysis of user preferences. However, the study overlooked the lesser-known attractions since the model only collected reviews from top attractions in the country.

The research [18] proposed a travel recommendation model to analyze the reviews collected from Google Maps and identify the most relevant locations for travellers based on the similarities and differences between the users' reviews. The Jaccard Similarity and Cosine Similarity were used to calculate the similarity scores. The algorithm ranked the most popular locations using a neural network and associated the users' preferences through the similarities of their reviews. On the other hand, [9] used Twitter data and built a system using a collaborative filtering framework with users' profile matrix and their interests. The travel-related tweets were mined for sentiment analysis, and a follow-up step was performed to determine the social media activity of their friends. The algorithm will generate travel recommendations and suggest various destinations based on relevant tweets. Unlike previous systems, the model was time-sensitive, allowing it to collect the users' most recent interests.

Moreover, [36] proposed a deep learning-based recommendation model to analyze the data from blogs, Google Maps and TripAdvisor to recommend travel activities in the country. Latent Dirichlet Allocation (LDA) was employed by the researchers for topic modelling in tourist blogs. These topics were used to extract the sentiments from Google and TripAdvisor reviews. The user history was extracted based on the information and a collaborative filtering technique was used to predict the most likely visited locations based on the users' preferences.

D. *Analyzing Geotagged Data*

The development of geotagging services and Web technologies have boosted the amount of geotagged data accessible. Through social media platforms like Foursquare, Facebook, and Flickr, individuals can now easily share their locations with others. Consequently, a growing corpus of research explored the use of geotagged data in personalized travel destination suggestions [37-39].

The authors in [38] presented a travel recommendation system that combines geotagged data with users' textual information. The multiclass SVM classifier was used to identify candidates from the user's travel history. The data was analyzed using a gradient-boosting regression model, which ranked the candidates based on their interests. Moreover, [19] proposed a weighted multi-information criteria matrix factorization model for recommending travel locations based on geotagged photos from Flickr. The model was built to examine the various aspects of a user's visit sequence, as well as the textual and visual information to recommend travel locations. The textual information in the photos was processed using Latent Dirichlet Allocation (LDA) to profile the attractions, and the model was tested on a sample of six Chinese cities.

The researchers in [37] combined the sequential and temporal information from the geotagged photos to build personalized itineraries based on the travel patterns of individual users. The model was developed using a collaborative filtering strategy to analyze the visit sequences and preferences of other users. In addition, [39] established a framework for determining the interests of Hong Kong tourists based on geotagged data. The study combined image processing, text processing, and clustering algorithms to evaluate geotagged data in three geographical locations, enabling the government to comprehend better and promote popular vacation spots.

Numerous studies have identified landmarks and tourist attractions based on geotagged data acquired using clustering algorithms [36-37], [39-40]. The geotagged data was clustered using several techniques, including K-means clustering, mean shift clustering, and density-based clustering to build a location database containing the travel records of users to different destinations.

E. *Summary*

According to the existing travel recommendation models, the most common technique used in developing travel recommendation models is the collaborative filtering approach, which involves analyzing the users' interactions or similarities. However, most studies have focused on leveraging reviews from travel websites or textual information for topic modelling from geotagged data. This demonstrates a deficiency in using the implicit information from the geotagged data, which does not require extra information. Additionally, clustering techniques are often used to group geotagged data to build a location database. Therefore, the proposed model would employ a clustering algorithm to identify locations from the geotagged data, and the collaborative approach will be utilized to compute user similarities. The model will then deliver personalized recommendations to the users based on their travel histories.

III. MATERIALS AND METHOD

The research process was structured as a sequence of stages designed to achieve the study's goals. The steps include data selection, data pre-processing, data transformation, model building, and evaluation. A flowchart is also provided to visualize the whole process as shown in Fig. 1.

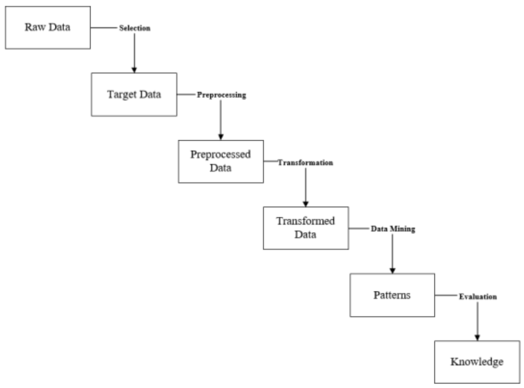


Fig. 1. KDD process.

A wide variety of data mining approaches are available for detecting patterns and interpreting data to develop a model. Knowledge Discovery in Databases (KDD) was chosen as the data mining methodology to develop the travel recommendation model. It is a process of studying data to uncover patterns that can be utilized to identify meaningful information [41]. Fig. 2 illustrates the entire process of developing the travel recommendation model.

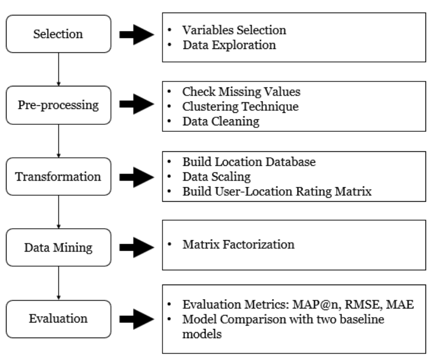


Fig. 2. Project workflow.

F. Selection

The literature review and research objectives provided a clear understanding to create the model. The selection procedure defined the target dataset for model building. The variables that could be used to construct the travel recommendation model were chosen at this stage.

1) Variables selection: As specified in the research scope, the dataset used for this study was obtained from Kaggle with 20,000 records and 13 attributes [21]. The attributes that contain the geographical information, the owner and the time were used to develop the recommendation model.

2) Dataset exploration: Data exploration is the process of examining data to comprehend better the data and reduces the likelihood of incorrect decisions [41]. The data properties were examined, and the geotagged data were visually analyzed. In addition, the missing values were also inspected.

G. Pre-Processing

Data pre-processing is typically performed to prepare the data before data modelling. The geotagged data were clustered, and redundant data was removed to improve the efficiency of the analysis.

1) Data cleaning: Data is the foundation of every data mining project. However, as data comes in a variety of formats and sizes, it must be thoroughly analyzed to ensure no discrepancies or outliers. Therefore, data cleaning was performed to identify missing values, outliers, and inconsistencies. The two most important attributes used in developing the travel recommendation model were the geotagged data, which includes latitude and longitude information. As a result, individuals with incomplete information for these two attributes were removed from the dataset, as the imputation of geographic location may disrupt the dataset's balance.

2) Clustering technique: The initial step in the development of the travel recommendations model was to identify the locations from the geotagged data. This process was carried out through the clustering technique, which was used to group the collected data into clusters. Three main types of clustering techniques were commonly used in this process: hierarchical clustering, partitional clustering, and density-based clustering [42].

Several studies have been conducted on identifying landmarks and hotspots from the geotagged data. A classical method used in discovering tourist attractions is the Density-based spatial clustering of applications with a noise clustering algorithm (DBSCAN) [37], [40], [43-45]. DBSCAN is effective in clustering geotagged data since it requires less knowledge to detect arbitrary shape clusters with varying densities. This method seems to be more effective when analyzing spatial data concerning latitude-longitude coordinates. The clustering process shall result in a dataset that resembles Table I.

TABLE I. DATASET AFTER DBSCAN

photo_id
user_id
lat
lon
Taken (time stamp)
location_id (cluster label)
cent_lat (cluster lat)
cent_lon (cluster lon)

3) Location database: A location database is built by applying the appropriate clustering technique which helped to label the geotagged data. The resulting database was found with the sorted locations visited according to the timestamp included. According to the study by [19], if the user concurrently uploads two geotagged posts during the same visit, the two posts should be regarded as one. This is to reduce the number of repeat visits by the same users and improve the quality of the location database. The duplicate records would be deleted if the time interval between two successive postings is less than three hours and assumed to have originated from the same visit.

Individuals with less than three geotagged posts were excluded from the dataset since the model was constructed using the collaborative approach. As illustrated in Table II, the final location database containing user travel histories shall have five variables: location id (cluster label), user id, lat, lon, and time taken (timestamp).

TABLE II. LOCATION DATABASE

location_id	user_id	lat	lon	time_taken

H. Transformation

The data transformation process involves arranging the data into a suitable form for modelling. In this stage, the user-location rating matrix was built using the pre-processed geotagged data.

1) User-location rating matrix: Using their historical travel records, individuals' preferences for a place may be inferred by the frequency of their visits to a travel destination. Inspired by the research of [19], the user-location rating table estimates the frequency of a user's visits to a travel location as ratings. Therefore, the number of times a user has visited a specific location is used to construct a new variable, ratings. As illustrated in Table III, three variables: user id, location id, and ratings depending on the frequency of visits were transformed into a matrix as shown in Table IV. The ratings were standardized using the min-max approach to ensure that the forecast is not skewed towards popular attractions.

TABLE III. USER-LOCATION RATING TABLE

location_id (cluster label)	User_id	Ratings

TABLE IV. USER-LOCATION RATING MATRIX

User_id	**location_id**			
	A	**B**	**C**	**D**
1	rating	rating	rating	rating
2	rating	rating	rating	rating
3	rating	rating	rating	rating

I. Data Mining

Data mining is the process of examining data to uncover hidden insights and patterns. This process aims to develop a personalized travel recommendation model that can provide the best possMatrix Factorization Model.

Multiple techniques are used in the construction of the recommendation model. According to the literature review, there are four main techniques such as content-based, collaborative-based, knowledge-based, and hybrid-based. This study used a collaborative approach based on user similarities since many prior studies relied on collaborative techniques to investigate user interactions and provide recommendations. For instance, [11] investigated destination reviews and user similarities to recommend vacation places, while [37] presented an itinerary planner by assessing different travel patterns from other users and matching the suggestions to the users' preferences.

TABLE V. USER-USER SIMILARITY MATRIX

User_id	**User_id**		
	1	**2**	**3**
1	similarity	similarity	similarity
2	similarity	similarity	similarity
3	similarity	similarity	similarity

Similarity value is one of the most crucial aspects when building a recommendation system using a collaborative method. The user-location rating matrix was used to generate the user-user similarity matrix (Table V) to construct the recommendation model. The cosine similarity algorithm was selected as the metric to determine the similarities between various users since it is one of the most extensively used and well-known similarity measures [18-19], [40], [46-50].

The recommendation model was built to predict the ratings to generate recommendations based on the user profiles from the user-user similarity matrix with a sparse rating matrix. Despite the popularity of the K-Nearest Neighbor (KNN) model as a collaborative-based technique, the KNN model required users to select the number of nearest neighbours, making the prediction unstable [51]. The study found that the non-negative matrix factorization model outperformed the k-nearest neighbour model in terms of accuracy and error metrics while constructing the movie rating recommendation system. Besides, [52] noted that the matrix factorization model was able to provide more precise pairwise preference scores and ranking predictions. Therefore, using the value generated from the cosine similarity algorithm, the matrix factorization model, which was extensively used in recommender systems in a variety of domains, was used to provide personalized recommendations to the target users [47], [53-55]. The system would be able to identify latent factors in the data and recommend the most appropriate destination according to their preferences without requiring additional features.

The matrix factorization can be equation as:

$$P\ x\ Q^{T} \approx R \qquad (1)$$

The main idea of a matrix factorization technique is to fit the rating matrix with a low-ranking approximation that considers the latent features. For instance, the matrix P in the equation represents the association between the user and its features. The matrix Q represents the association between the item and its features. The prediction of the rating is the dot product of the latent factors. The model is fueled by the ratings provided by the user-location rating matrix. The prediction values will be used to rank the top-n suggestions. This aligns with the project's goal of providing personalized suggestions ranked according to the prediction values.

J. Evaluation

Model evaluation is an integral part of the data mining process for measuring the performance of models using a variety of evaluation metrics. Multiple studies indicate that Mean Average Precision (MAP), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are the common assessment metrics for measuring the performance of a recommendation model [19], [30], [38], [47], [53], [55-57].

1) Mean average precision@n (MAP@n) [38]

The formula:

$$Precision = \frac{True\ Positive}{Total\ Postitive\ Results} \quad (2)$$

$$MAP@n = \frac{1}{N}\sum_{i=1}^{N_q} AP_i, AP_i = \frac{1}{r} \quad (3)$$

2) Mean absolute error (MAE)

The formula:

$$MAE = \frac{1}{N}\sum |Predicted\ Ratings - Actual\ Ratings| \quad (4)$$

3) Root mean square error (RMSE)

The formula:

$$RMSE = \sqrt{\sum \frac{(Predicted\ Ratings - Actual\ Ratings)^2}{N}} \quad (5)$$

4) Comparison with two baseline methods: The performance of the proposed travel recommendation model was compared with two baseline models. The random selection technique and the popularity-based strategy were chosen for comparative studies [37], [47], [58]. The popularity-based technique recommends the most popular vacation destination based on an overall popularity score. The random selection strategy, on the other hand, generates travel destinations at random from the location database, ignoring similarities between users.

The data was split into training (80%) and testing (20%). The recommendations were made based on the users' past travel experiences, and the recommended locations were ranked based on the projected values. The top-n recommendations to the target users were compared with the actual ratings.

IV. Implementation

A. Data Pre-Processing

The data pre-processing is crucial to building the model since it allows the dataset to be prepared for modelling purposes.

The data was obtained from Kaggle with over 20,000 records and 13 attributes as described in Table VI.

1) Variables selection: The project's objective was to recommend travel destinations to users based on geotagged information collected. As shown in Table VII, the study employed only five variables for building the model such as picture id, user id, geographical information, including the latitude and longitude of the images, and the time at which the photos were taken.

TABLE VI. Dataset Description before Variables Selection

No	Attributes	Description
1	photo_id	photo id
2	owner	user id related to the owner of the photo
3	gender	owner's gender
4	occupation	occupation of owner
5	title	title of photo
6	description	description of photo
7	tags	photo tag
8	faves	photo's favorite rate
9	lat	photo's latitude
10	lon	photo's longitude
11	u_city	owner's city
12	u_country	owner's country
13	taken	the time of the photo taken

2) Data exploration: The data type of the dataset was inspected after eliminating unnecessary attributes. Suitable type conversions were done to the variables such as date, latitude, and longitude to prevent slower operations during data transformation and model construction. The dataset was examined for missing values and confirmed with no imputation or removal of data required.

3) Clustering algorithm: The initial step in developing the travel recommendation model involved identifying the locations of the geotagged data. The study [59] highlighted that it is often challenging to process spatial data due to the existence of redundant points. By transforming the number of latitude-longitude coordinates into the corresponding values generated by the clustering technique, DBSCAN can reduce the size of a geographical data set to a small collection of representative points. The data were grouped into clusters to serve as a location for recommendations. The latitude and longitude data were extracted as dbscan_data as the first step.

The two main parameters for the DBSCAN algorithm are the epsilon (eps) and the minimum points (MinPts). The epsilon specifies the radius of a neighbourhood around the center point of the clusters, and it is important to determine the optimal number of clusters. If the eps value is too low, a significant amount of the data will be omitted from the cluster. This is because the value is insufficient to produce a dense region. Conversely, if the value is very high, many objects will be merged into a cluster, making the clustering meaningless. Besides, the parameter MinPts specifies the minimum number of points necessary to create a cluster. The estimation of the various parameters used is often a challenge during the development of an algorithm. Therefore, different combinations of eps and MinPts values were examined to discover the optimal values.

Research [59] stated the haversine metric was used as the metric of the DBSCAN algorithm to minimize the noise generated by the random selection process by computing the great-circle distances between the various points in the data set. Given their respective latitudes and longitudes, the haversine formula relates the great-circle distance of a sphere to two locations on a specified plane. The parameter and coordinates were then converted to radians to ensure the algorithm to perform precise calculations.

Fig. 3 shows the number of clusters generated by various parameter combinations. The optimal value for eps and MinPts was determined using the elbow approach, which is a basic procedure used in cluster analysis [60]. As a result, the eps and minimum points selected were 0.15 and 10, respectively.

The geographical coordinates were converted from 20,000 data points to 181 clusters using the given parameters. With cluster_num = 0 as the noises in the data, 180 clusters were found as the representative points of the travel locations for recommendations. The coordinates of the geotagged data that were formed as part of the cluster were labelled by its cluster labels using the points closest to the cluster's centroid. It was accomplished by taking a set of points and returning to the centermost point of the cluster.

Fig. 4 illustrates the original data set which was reduced to a cluster-representative collection of points where different colours correspond to each cluster formed by the DBSCAN algorithm. The grey dots represent the outliers of the geotagged data points, often known as the dataset's noise.

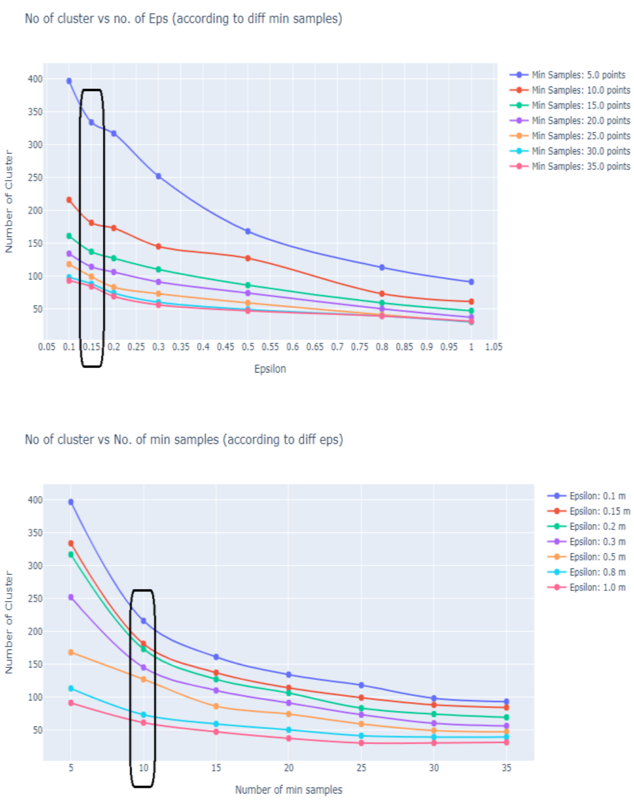


Fig. 3. Selecting DBSCAN parameters.

TABLE VII. DATASET DESCRIPTION AFTER VARIABLES SELECTION

No	Attributes	Description
1	photo_id	photo id
2	owner	user id related to the owner of the photo
3	lat	photo's latitude
4	lon	photo's longitude
5	taken	the time of the photo taken

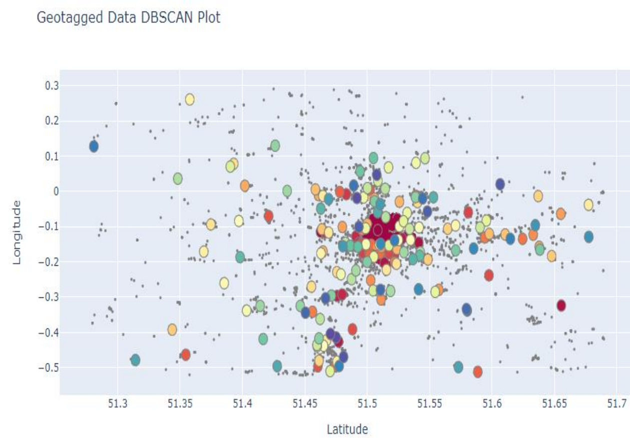


Fig. 4. DBSCAN result.

4) Remove outliers: The outliers were eliminated as found as noise in the dataset and not beneficial for the recommendation process after constructing the cluster labels for each geotagged data set.

5) Remove duplicate data: It is important to note that users may take multiple photos of the same location while visiting a venue. Therefore, if the timestamps between two photos taken at the same location are less than three hours, the visit must be treated as a single visit. The most recent timestamps of the images taken at the same place are then used to determine the time. The elimination of duplicate data was done by merging users with visits within three hours. In addition, for consistency, a random number between 100 and 999 was generated in place of the cluster labels 0 to 180.

6) Remove users with less than three visits: Users who have visited less than three distinct places were eliminated as the model was developed using a collaborative filtering approach.

B. Data Transformation

The rating matrix to construct the recommendation model was formed by calculating the number of times a user visited a certain location as a rating. The ratings were scaled to a range between 1 and 5 using the min-max scale function to eliminate bias in the training phase and enhance the efficiency of the data mining process. After the data scaling process was completed, the user-location rating matrix was generated using the pivot_table function to develop a travel location recommendation model and provide users with ideal suggestions based on their previous visits.

The final dataset was split into two: training (80%) and testing (20%) to assess the recommendation model.

C. Data Mining

This section discusses the matrix factorization model, a widely used technique in the collaborative filtering approach to construct the travel location recommendation model.

1) Matrix factorization model: The rating matrix is sparse by nature. The goal of the modelling process was to forecast the ratings of the areas that have not been visited by the target user through user similarities. Therefore, the user-user similarity matrix is crucial for determining the possibility that they will visit other travel destinations. According to [61], the sparsity in the rating matrix is a major factor that affects the performance of collaborative filtering systems, and the matrix factorization model is effective in addressing the insufficiency of ratings. Therefore, the modelling process utilized the matrix factorization model to factorize the various similarities between different pairs of users. The main objective of this process was to predict the missing values of the user-user similarity matrix. The cosine similarity was selected as the metric to predict the similarities between the users. The matrix factorization model decomposed the original matrix of user preferences into two smaller matrix elements, known as latent factors. The model discovered the hidden features of the interactions between different users and analysed the various factors that affect the users' behaviour to recommend the most appropriate destination according to their preferences. The approach was inspired by the study by [57], who revealed that the matrix factorization performed well with sparse data using movie ratings.

After determining various similarities between every pair of users, a weighted average of the ratings from the users like the target user was then used to calculate the ratings to a location for the current user. The projected rating of a specific location was the weighted sum of the ratings given to a certain location by the number of users like the target user. The predicted rating was the expected value that the target users will be assigned to the specific location.

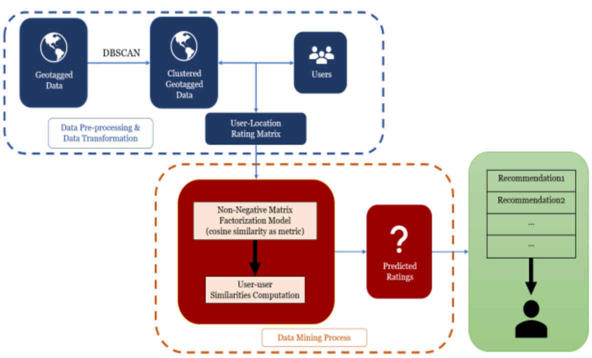


Fig. 5. Data mining process.

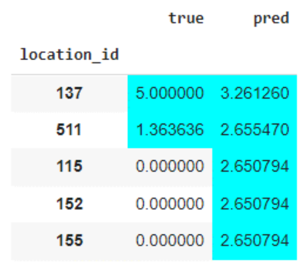
Fig. 5 provides a summary of the model implementation. Multiple approaches were used to prepare and transform the data, and the data mining process enables the model to learn and predict ratings for the target user to a certain location.

V. Results and Discussion

This Section provides a comprehensive analysis of the performance of the developed model. Several evaluation metrics were used to test the model's performance. The results of the study were analyzed through a comparative analysis with two baseline models.

A. Results

The testing datasets were used to construct the test rating matrix for evaluating the proposed model. In the testing dataset, the ratings were also scaled to a range of 1 to 5. Using the proposed model, the predict function was defined to forecast the ratings of locations for a user. Fig. 6 shows the top five suggestions for the representative target user, 41087279@N00. The ratings are displayed side by side with the predictions.



location_id	true	pred
137	5.000000	3.261260
511	1.363636	2.655470
115	0.000000	2.650794
152	0.000000	2.650794
155	0.000000	2.650794

Fig. 6. Prediction result.

The proposed model was evaluated against two different baseline approaches selected from past studies, and the assessment was conducted using the same dataset [37], [47], [58]. One of these is the popularity-based method, which used a general popularity score as the basis for its recommendations. It considered the number of unique visits to these locations. On the other hand, the random model generated random travel destinations for the target users regardless of their similarities or popularity scores.

B. Model Comparison

1) Mean average precision@n (MAP@n): The mean average precision (MAP) is a measure that takes into the list of recommendations and compares it with the true set. The n represents the number of recommendations generated to the users. Using n = 5, 10, 15, the MAP for each model was calculated and tabulated for comparison. Based on Fig. 7 and Table VIII, the results show that the proposed matrix factorization model got an average precision value of 0.83, which is higher than the popularity and random models.

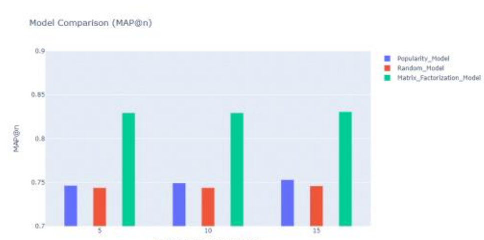


Fig. 7. MAP@n.

TABLE VIII. MAP@N

MAP@n	Popularity Model	Random Model	Matrix Factorization Model
5	0.746	0.744	0.829
10	0.749	0.744	0.829
15	0.753	0.745	0.831

2) Root mean square error (RMSE): The Root Mean Square Error (RMSE) is a statistical tool used to assess the accuracy of rating predictions. It measures the square root of the difference between predicted and actual values. As shown in Fig. 8 and Table IX, the matrix factorization model got an RMSE value of 1.36, which is the lowest compared to the two baseline models.

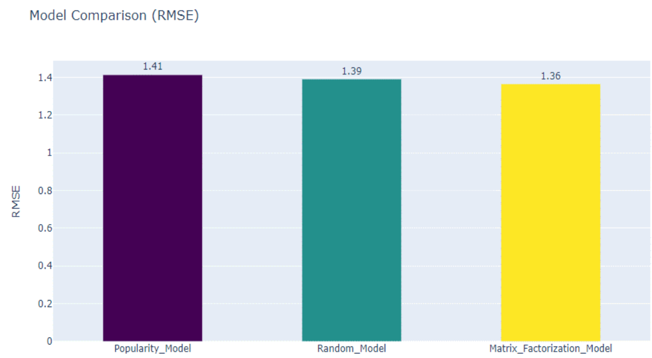


Fig. 8. RMSE.

TABLE IX. RMSE

Models	RMSE
Popularity Model	1.41
Random Model	1.39
Matrix Factorization Model	1.36

3) Mean absolute error (MAE): The Mean Absolute Error (MAE) represents the deviations between the model's predictions and the actual results. Fig. 9 and Table X reveal that the proposed matrix factorization model got the lowest MAE of 1.24 value among the three models. In contrast, the MAE value for the popularity and random models is 1.27 and 1.28, respectively.

The Mean Average Precision@n (MAP@n), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) of the three models are summarized in Table XI. The results demonstrate that the proposed model has the lowest RMSE and MAE values and the highest MAP@n regardless of the number of recommendations. This proves that the user-based collaborative filtering technique, which computes user-user similarities using a matrix factorization model, effectively identifies the ideal destination for the target users based on their previous visits.

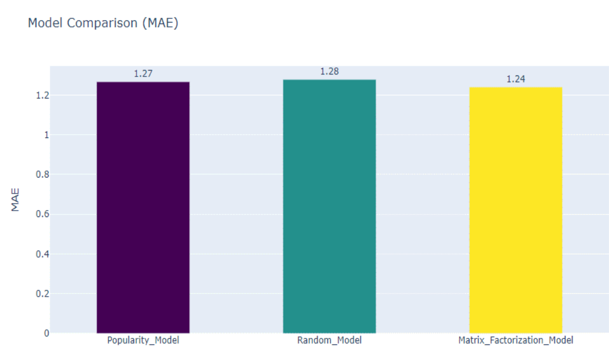


Fig. 9. MAE.

TABLE X. MAE

Models	MAE
Popularity Model	1.27
Random Model	1.28
Matrix Factorization Model	1.24

TABLE XI. MODEL COMPARISON

Models	MAP@n			RMSE	MAE
	5	10	15		
Popularity Model	0.746	0.749	0.753	1.41	1.27
Random Model	0.744	0.744	0.745	1.39	1.28
Matrix Factorization Model (Proposed Model)	0.829	0.829	0.831	1.36	1.24

VI. CONCLUSION

This Section summarizes the findings and the study's contribution to the research field. Limitations and future recommendations are additionally highlighted to improve the algorithm's performance and enhance its usability for the general population.

A. Conclusion

The rapid development and growth of the tourism industry has led to the need for more effective tools and methods to help travellers make informed decisions when planning their trips. Numerous studies have been conducted on the development of effective travel recommendation systems, but most of them have been reliant on reviews and descriptions of the attractions.

The study develops a recommendation system for travellers using geotagged data to provide personalized location recommendations based on user interactions. The evaluation metrics used, such as MAP@n, RMSE and MAE, revealed that the proposed model outperforms the two baseline models chosen by exhibiting recommendations with the highest MAP values and the lowest RMSE and MAE values. As per the findings, the proposed model obtained the highest MAP value using the different number of recommendations generated, with an average value of 0.83. Further, compared to the two baseline models, the proposed model got the lowest RMSE and MAE, with values of 1.36 and 1.24, respectively. This proves that the matrix factorization model effectively generates personalized location recommendations based on users' past visits and interactions with other users.

The study has significant implications for the tourism industry, as the proposed system can help travellers make informed decisions when planning their trips. The use of geotagged data provides a more comprehensive and unbiased view of travel destinations, as it considers both popular and less popular spots. Additionally, the results of this study also contribute to the existing literature on travel recommendation systems by showcasing the efficacy of utilizing geotagged data to generate personalized recommendations.

B. Contributions and Importance of the Study

Overall, the study has accomplished its aim of analyzing the impact of geotagged data in selecting attractions and proposing a travel recommendation model based on geotagged data. The proposed travel recommendation model can analyze the collective behaviour of tourists and identify regions that are ideal for them. It also introduces serendipity by enabling users to discover new interests in different areas depending on the interests indicated by other similar users based on the data collected.

In conclusion, the proposed model has the potential to be a valuable tool for users who have uploaded geotagged social media posts to get travel destination ideas from other places. This can ultimately reduce the time spent browsing through different websites to find the ideal destination to travel to. Additionally, the tourism sector may incorporate this model into their applications to promote tourism in their respective countries to create revenue and contribute to their Gross Domestic Product (GDP). With further implementation in the future, this can provide significant benefits to both users and the tourism industry.

C. Limitations and Future Recommendations

The collaborative filtering technique does not require domain knowledge since embeddings are automatically learnt, and the matrix factorization can solve the sparsity problem. However, the proposed model suffers from the cold-start problem, which occurs for users that are relatively new due to insufficient connection with other users. To address this issue, users with less than three visits were excluded from the recommendation process. It is recommended to incorporate other algorithms such as content-based filtering to build a hybrid recommender to eliminate the cold-start issue. Content-based filtering can be used to construct user profiles by collecting user information based on their social media postings or through a questionnaire.

It is also recommended to build a mobile application or a graphical user interface (GUI) using the developed recommendation model to provide personalized recommendations. The application should have an interactive interface that allows the users to select their previous visits and display the various locations that it recommends. In addition to providing locations as recommendations, the proposed model should also add side features collected from sites like Google Maps to enhance the location profiles. For example, it can provide the type of activities and opening hours in the area to allow users to understand more about the recommended locations.

REFERENCES

[1] World Travel & Tourism Council, “Economic Impact Reports,” 2020. https://wttc.org/Research/Economic-Impact (accessed Oct. 19, 2021).

[2] V. Kazandzhieva and H. Santana, “E-tourism: Definition, development and conceptual framework,” Tourism: An International Interdisciplinary Journal, vol. 67, no. 4, pp. 332–350, 2019.

[3] S. Singh and A. Bashar, “A bibliometric review on the development in e-tourism research,” International Hospitality Review, vol. ahead-of-print, no. ahead-of-print, Jan. 2021, doi: 10.1108/IHR-03-2021-0015.

[4] O. Wyman, “To Recovery & Beyond | Future of Travel & Tourism | World Travel & Tourism Council (WTTC),” 2020. Accessed: Oct. 10, 2021. [Online]. Available: https://wttc.org/Initiatives/To-Recovery-Beyond

[5] X. Liu, F. Mehraliyev, C. Liu, and M. Schuckert, “The roles of social media in tourists’ choices of travel components,” Tour Stud, vol. 20, no. 1, pp. 27–48, Apr. 2019, doi: 10.1177/1468797619873107.

[6] B. Hysa, A. Karasek, and I. Zdonek, “Social Media Usage by Different Generations as a Tool for Sustainable Tourism Marketing in Society 5.0 Idea,” Sustainability, vol. 13, no. 3, 2021, doi: 10.3390/su13031018.

[7] B. T. Khoa, N. M. Ly, V. T. T. Uyen, N. T. T. Oanh, and B. T. Long, “The impact of Social Media Marketing on the Travel Intention of Z Travelers,” in 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1–6. doi: 10.1109/IEMTRONICS52119.2021.9422610.

[8] S. W. Litvin, R. E. Goldsmith, and B. Pan, “A retrospective view of electronic word-of-mouth in hospitality and tourism management,” International Journal of Contemporary Hospitality Management, vol. 30, no. 1, pp. 313–325, Jan. 2018, doi: 10.1108/IJCHM-08-2016-0461.

[9] P. Nitu, J. Coelho, and P. Madiraju, “Improvising personalized travel recommendation system with recency effects,” Big Data Mining and Analytics, vol. 4, no. 3, pp. 139–154, 2021, doi: 10.26599/BDMA.2020.9020026.

[10] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, “Tourism recommendation system based on semantic clustering and sentiment analysis,” Expert Syst Appl, vol. 167, p. 114324, Apr. 2021, doi: 10.1016/J.ESWA.2020.114324.

[11] I. Y. Choi, Y. U. Ryu, and J. K. Kim, “A recommender system based on personal constraints for smart tourism city,” Asia Pacific Journal of Tourism Research, vol. 26, no. 4, pp. 440–453, Apr. 2021, doi: 10.1080/10941665.2019.1592765.

[12] J.-S. Kim et al., “Location-Based Social Network Data Generation Based on Patterns of Life,” in 2020 21st IEEE International Conference on Mobile Data Management (MDM), 2020, pp. 158–167. doi: 10.1109/MDM48529.2020.00038.

[13] X. Wei, Y. Qian, C. Sun, J. Sun, and Y. Liu, “A survey of location-based social networks: problems, methods, and future research directions,” Geoinformatica, 2021, doi: 10.1007/s10707-021-00450-1.

[14] A. Hardy, “Tracking via Geotagged Social Media Data,” in Tracking Tourists, Goodfellow Publishers, 2020. doi: 10.23912/9781911635383-4575.

[15] C. Barros, J. Gutiérrez, and J. García-Palomares, “Geotagged data from social media in visitor monitoring of protected areas; a scoping review,” Current Issues in Tourism, pp. 1–17, May 2021, doi: 10.1080/13683500.2021.1931053.

[16] V. Taecharungroj and B. Mathayomchan, “Traveller-generated destination image: Analysing Flickr photos of 193 countries worldwide,” International Journal of Tourism Research, vol. 23, no. 3, pp. 417–441, Apr. 2021, doi: 10.1002/JTR.2415.

[17] M.-I. Ana and L.-G. Istudor, “The Role of Social Media and User-Generated-Content in Millennials’ Travel Behavior,” Management Dynamics in the Knowledge Economy, vol. 7, no. 1, pp. 87–104, 2019, doi: http://dx.doi.org/10.25019/MDKE/7.1.05.

[18] S. Bairavel and M. Krishnamurthy, “User preference and reviews analysis with neural networks for travel recommender systems,” International Journal of Engineering Research and Technology, vol. 13, no. 8, pp. 1896–1900, 2020, doi: 10.37624/IJERT/13.8.2020.1896-1900.

[19] D. Lyu, L. Chen, Z. Xu, and S. Yu, “Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geo-tagged photos,” Applied Intelligence, vol. 50, no. 3, pp. 924–938, 2020, doi: 10.1007/s10489-019-01566-6.

[20] S.-T. Park and C. Liu, “A study on topic models using LDA and Word2Vec in travel route recommendation: focus on convergence travel and tours reviews,” Pers Ubiquitous Comput, 2020, doi: 10.1007/s00779-020-01476-2.

[21] A. Sariaslani, “Flickr (London City),” 2020. https://www.kaggle.com/datasets/amiralisa/flickr_london (accessed Apr. 04, 2022).

[22] WHO, “WHO Director-General’s opening remarks at the media briefing on COVID-19 - 11 March 2020,” World Helath Organization, 2020. https://www.who.int/director-general/speeches/detail/who-director-

general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed Dec. 05, 2021).

[23] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in Aaai/iaai, 1998, pp. 714–720.

[24] A. Ansari, S. Essegaier, and R. Kohli, "Internet Recommendation Systems," Journal of Marketing Research, vol. 37, no. 3, pp. 363–375, 2000, doi: 10.1509/jmkr.37.3.363.18779.

[25] C. C. Aggarwal, "An Introduction to Recommender Systems," Recommender Systems: The Textbook. Springer International Publishing, pp. 1–28, 2016. doi: 10.1007/978-3-319-29659-3_1.

[26] Z.-P. Zhang, Y. Kudo, T. Murai, and Y.-G. Ren, "Enhancing Recommendation Accuracy of Item-Based Collaborative Filtering via Item-Variance Weighting," Applied Sciences, vol. 9, no. 9, 2019, doi: 10.3390/app9091928.

[27] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, Recommender systems: an introduction. Cambridge University Press, 2010.

[28] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: A survey," Decis Support Syst, vol. 74, pp. 12–32, 2015, doi: https://doi.org/10.1016/j.dss.2015.03.008.

[29] B. Walek and V. Fojtik, "A hybrid recommender system for recommending relevant movies using an expert system," Expert Syst Appl, vol. 158, p. 113452, 2020, doi: https://doi.org/10.1016/j.eswa.2020.113452.

[30] S. Khalifeh and A. A. Al-Mousa, "A Book Recommender System Using Collaborative Filtering Method," International Conference on Data Science, E-Learning and Information Systems 2021, pp. 131–135, 2021, doi: 10.1145/3460620.3460744.

[31] J. K. Tarus, Z. Niu, and G. Mustafa, "Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning," Artif Intell Rev, vol. 50, no. 1, pp. 21–48, 2018, doi: 10.1007/s10462-017-9539-5.

[32] Y. Deldjoo, M. Schedl, and M. Elahi, "Movie Genome Recommender: A Novel Recommender System Based on Multimedia Content," in 2019 International Conference on Content-Based Multimedia Indexing (CBMI), 2019, pp. 1–4. doi: 10.1109/CBMI.2019.8877452.

[33] Adiyansjah, A. A. S. Gunawan, and D. Suhartono, "Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks," Procedia Comput Sci, vol. 157, pp. 99–109, 2019, doi: https://doi.org/10.1016/j.procs.2019.08.146.

[34] S. Ali, Y. Hafeez, M. Humayun, N. S. M. Jamail, M. Aqib, and A. Nawaz, "Enabling recommendation system architecture in virtualized environment for e-learning," Egyptian Informatics Journal, vol. 23, no. 1, pp. 33–45, 2022, doi: https://doi.org/10.1016/j.eij.2021.05.003.

[35] 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," Pers Ubiquitous Comput, vol. 23, no. 2, pp. 181–197, 2019, doi: 10.1007/s00779-018-01194-w.

[36] W. Shafqat and Y.-C. Byun, "A Context-Aware Location Recommendation System for Tourists Using Hierarchical LSTM Model," Sustainability, vol. 12, no. 10, 2020, doi: 10.3390/su12104107.

[37] G. Cai, K. Lee, and I. Lee, "Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos," Expert Syst Appl, vol. 94, pp. 32–40, Apr. 2018, doi: 10.1016/J.ESWA.2017.10.049.

[38] X. Sun, Z. Huang, X. Peng, Y. Chen, and Y. Liu, "Building a model-based personalised recommendation approach for tourist attractions from geotagged social media data," Int J Digit Earth, vol. 12, no. 6, pp. 661–678, Apr. 2019, doi: 10.1080/17538947.2018.1471104.

[39] L. Zhong, L. Yang, J. Rong, and H. Kong, "A Big Data Framework to Identify Tourist Interests Based on Geotagged Travel Photos," IEEE Access, vol. 8, pp. 85294–85308, 2020, doi: 10.1109/ACCESS.2020.2990949.

[40] H. Huang, "Context-Aware Location Recommendation Using Geotagged Photos in Social Media," ISPRS International Journal of Geo-Information 2016, Vol. 5, Page 195, vol. 5, no. 11, p. 195, Apr. 2016, doi: 10.3390/IJGI5110195.

[41] S. A. Mohd Selamat, S. Prakoonwit, R. Sahandi, W. Khan, and M. Ramachandran, "Big data analytics—A review of data-mining models for small and medium enterprises in the transportation sector," Wiley Interdiscip Rev Data Min Knowl Discov, vol. 8, no. 3, p. e1238, May 2018, doi: 10.1002/WIDM.1238.

[42] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," in Emerging Technology in Modelling and Graphics, 2020, pp. 69–83.

[43] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen, "Travel Recommendation Using Geo-tagged Photos in Social Media for Tourist," Wirel Pers Commun, vol. 80, no. 4, pp. 1347–1362, 2015, doi: 10.1007/s11277-014-2082-7.

[44] W. Höpken, M. Müller, M. Fuchs, and M. Lexhagen, "Flickr data for analysing tourists' spatial behaviour and movement patterns," Journal of Hospitality and Tourism Technology, vol. 11, no. 1, pp. 69–82, Apr. 2020, doi: 10.1108/JHTT-08-2017-0059.

[45] K. N. S. Behara, A. Bhaskar, and E. Chung, "A DBSCAN-based framework to mine travel patterns from origin-destination matrices: Proof-of-concept on proxy static OD from Brisbane," Transp Res Part C Emerg Technol, vol. 131, p. 103370, 2021, doi: https://doi.org/10.1016/j.trc.2021.103370.

[46] M. Grbovic et al., "E-commerce in Your Inbox: Product Recommendations at Scale," Jun. 2016, doi: 10.1145/2783258.2788627.

[47] Z. Xu, L. Chen, Y. Dai, and G. Chen, "A Dynamic Topic Model and Matrix Factorization-Based Travel Recommendation Method Exploiting Ubiquitous Data," IEEE Trans Multimedia, vol. 19, no. 8, pp. 1933–1945, 2017, doi: 10.1109/TMM.2017.2688928.

[48] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, vol. 16, no. 3, pp. 261–273, 2015, doi: https://doi.org/10.1016/j.eij.2015.06.005.

[49] Z. Ghaemi and M. Farnaghi, "A Varied Density-based Clustering Approach for Event Detection from Heterogeneous Twitter Data," ISPRS Int J Geoinf, vol. 8, no. 2, 2019, doi: 10.3390/ijgi8020082.

[50] W. Yin, Y. Sun, and J. Zhao, "Personalized Tourism Route Recommendation System Based on Dynamic Clustering of User Groups," in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 1148–1151. doi: 10.1109/IPEC51340.2021.9421158.

[51] E. Torstensson, "Comparing neighborhood and matrix factorization models in recommendation systems: Saving the user some clicks," 2019.

[52] S. Kalloori, F. Ricci, and M. Tkalcic, "Pairwise Preferences Based Matrix Factorization and Nearest Neighbor Recommendation Techniques," in Proceedings of the 10th ACM Conference on Recommender Systems, 2016, pp. 143–146. doi: 10.1145/2959100.2959142.

[53] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," Inf Process Manag, vol. 54, no. 3, pp. 463–474, 2018, doi: https://doi.org/10.1016/j.ipm.2018.02.005.

[54] S. Bin and G. Sun, "Matrix Factorization Recommendation Algorithm Based on Multiple Social Relationships," Math Probl Eng, vol. 2021, p. 6610645, 2021, doi: 10.1155/2021/6610645.

[55] R. Wang, H. K. Cheng, Y. Jiang, and J. Lou, "A novel matrix factorization model for recommendation with LOD-based semantic similarity measure," Expert Syst Appl, vol. 123, pp. 70–81, 2019, doi: https://doi.org/10.1016/j.eswa.2019.01.036.

[56] C. C. Aggarwal, "Evaluating recommender systems," in Recommender systems, Springer, 2016, pp. 225–254.

[57] P. Pirasteh, D. Hwang, and J. J. Jung, "Exploiting matrix factorization to asymmetric user similarities in recommendation systems," Knowl Based Syst, vol. 83, pp. 51–57, 2015, doi: https://doi.org/10.1016/j.knosys.2015.03.006.

[58] K. H. Lim, "Recommending Tours and Places-of-Interest Based on User Interests from Geo-Tagged Photos," in Proceedings of the 2015 ACM SIGMOD on PhD Symposium, 2015, pp. 33–38. doi: 10.1145/2744680.2744693.

[59] G. Boeing, "Clustering to Reduce Spatial Data Set Size," SSRN Electronic Journal, Mar. 2018, doi: 10.48550/arxiv.1803.08101.

[60] T. Luo, X. Zheng, G. Xu, K. Fu, and W. Ren, "An Improved DBSCAN Algorithm to Detect Stops in Individual Trajectories," ISPRS Int J Geoinf, vol. 6, no. 3, 2017, doi: 10.3390/ijgi6030063.

[61] R. Duan, C. Jiang, and H. K. Jain, "Combining review-based collaborative filtering and matrix factorization: A solution to rating's sparsity problem," Decis Support Syst, vol. 156, p. 113748, 2022, doi: https://doi.org/10.1016/j.dss.2022.113748.