






## Research Article

# A Hybrid Method to Solve Data Sparsity in Travel Recommendation Agents Using Fuzzy Logic Approach

Mehrbakhsh Nilashi <sup>1,2</sup>, Rabab Ali Abumalloh <sup>3</sup>, Mesfer Alrizq<sup>4</sup>, Ahmed Almulihi,<sup>5</sup>  
O. A. Alghamdi,<sup>6</sup> Murtaza Farooque <sup>7</sup>, Sarminah Samad <sup>8</sup>, Saidatulakmal Mohd,<sup>2</sup>  
and Hossein Ahmadi <sup>9</sup>

<sup>1</sup>UCSI Graduate Business School, UCSI University, No. 1 Jalan Menara Gading, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia

<sup>2</sup>Centre for Global Sustainability Studies (CGSS), Universiti Sains Malaysia, Penang, Malaysia

<sup>3</sup>Computer Department, Applied College, Imam Abdulrahman Bin Faisal University, P.O. Box. 1982, Dammam, Saudi Arabia

<sup>4</sup>Information Systems Department, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia

<sup>5</sup>Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

<sup>6</sup>Business Administration Dept, Applied College, Najran University, Najran, Saudi Arabia

<sup>7</sup>Department of MIS, Dhofar University, Salalah, Oman

<sup>8</sup>Department of Business Administration, College of Business and Administration, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

<sup>9</sup>Centre for Health Technology, Faculty of Health, University of Plymouth, Plymouth PL4 8AA, UK

Correspondence should be addressed to Mehrbakhsh Nilashi; nilashidotnet@hotmail.com

Received 26 December 2021; Revised 3 January 2022; Accepted 9 June 2022; Published 27 June 2022

Academic Editor: Heng Liu

Copyright © 2022 Mehrbakhsh Nilashi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Travel recommendation agents have been a helpful tool for travelers in their decision-making for destination choices. It has been shown that sparsity can significantly impact on the accuracy of recommendation agents. The COVID-19 outbreak has affected the tourism and hospitality industry of almost all countries in the world. Tourists who have planned to travel are canceling or postponing trips due to this pandemic. Accordingly, this will impact the rate of travelers' online reviews on tourism products. Hence, the lack of data, in terms of ratings and textual reviews on hotels, will be a major issue for travel recommendation agents during the COVID-19 outbreak in the context of tourism and hospitality. This will be a new challenge for the researchers in the development of travel recommendation agents. Machine learning has been found to be effective in dealing with the data sparsity in recommendation agents. Therefore, developing new algorithms would be helpful to overcome the sparsity issue in travel recommendation agents. This research provides a new method through neurofuzzy, dimensionality reduction, and clustering techniques and evaluates it on the TripAdvisor dataset to see its effectiveness in solving the sparsity issue. The results showed that the method which used the fuzzy logic technique with the aid of clustering, dimensionality reduction, and fuzzy logic is more efficient in addressing the sparsity problem and presenting more accurate results. The results of the method evaluation are presented and discussed, and several suggestions are provided for future studies.

## 1. Introduction

Recommendation systems are software agents that aim to solve the information overload problem and enable real-time decision-making. The use of recommendation agents in

tourism has been effective for travelers' decision-making [1–3]. These systems have helped travelers to find the most suitable destinations according to their choice preferences [4]. It was shown that the accuracy of the recommendations relies on the richness of available data [5]. In fact, data

quality plays a vital part in recommending accurate items to the users. Sparsity which is a major issue in these systems [6–11] has significantly impacted the accuracy of the recommendation in many domains of application such as healthcare, tourism, and e-commerce.

The COVID-19 outbreak has affected the tourism and hospitality industry of almost all countries in the world [12, 13]. Tourists who have planned to travel are canceling or postponing trips due to this pandemic. This has negatively impacted the hospitality industry. Thus, many hotel managers will not receive enough feedback from travelers to measure the quality of the services. In fact, there will not be enough online customers' reviews on the performance criteria of hotels for their evaluation. Thus, the lack of data, in terms of quantitative (numerical ratings) and qualitative (textual reviews) ones on hotels, will be a major issue for travel recommendation agents during the COVID-19 outbreak, particularly considering the tourism and hospitality sector. It is therefore important to provide appropriate strategies to obtain enough and updated data regarding the tourism products from the users to have accurate travel recommendation agents during and in the post-COVID-19 era. In addition, it is vital to overcome the sparsity issue of user-hotel interaction data to have recommendation agents with reliable recommendations.

To address real-world obstacles, there is a need to handle several uncertain variables. Referring to the changing environment of uncertainty, several shortcomings in data such as fuzziness, randomness, incompleteness, and indistinguishability can exist [14]. Fuzzy logic presents a broad range of approaches to inspect the uncertainty in data [15]. Contrary to traditional set theory, in which items are either classified into a group or not, in fuzzy set theory, items can be a part of a group to some degree [16]. This theory has been deployed to represent qualitative data effectively and presented the ability to address several issues with very robust outcomes [17]. The use of fuzzy logic in previous literature has been explored in several fields like software project management [18], electronic learning [19], and the trust model [20]. Recommender Systems (RSs) based on fuzzy logic have been advanced since 2008 [21]. Hence, several researchers have explored fuzzy logic in several domains related to recommender systems like consensus ranking [22], item and trust-aware collaborative filtering [23], correlation based similarity [24], competence RSs [25], situation-aware collaborative filtering (CF) [26], tourism system [27], stock market [28], movie RSs [29], automatic group RSs [30], knowledge-based RSs [31], and multicriteria collaborative filtering [32].

This research provides a new method through fuzzy logic, dimensionality reduction, and clustering techniques and evaluates it on the TripAdvisor dataset to see its effectiveness in solving the sparsity issue. Data segmentation was deployed using Expectation-Maximization (EM), in which similar users were better detected with lower computation time. EM was deployed for different numbers of clusters and the best cluster size was considered for the next stage. Dimensionality reduction was applied using Higher-Order Singular Value Decomposition (HOSVD). We also

use adaptive neurofuzzy inference system (ANFIS) to predict users' overall ratings according to the predicted criteria ratings. Finally, the recommendations were presented using a dense tensor of the data. The findings of the study were evaluated using several measures. The results of the method evaluation are presented and discussed, and several suggestions are provided for future studies.

The remainder of this paper is structured as follows. In Section 2, we present a bibliometric analysis of the sparsity problem in the recommender system literature. In Section 3, we provide a literature review on RSs in tourism and hospitality. In Section 4, we provide the proposed method and the mathematical background of the proposed techniques. In Section 5, the data collection is provided. Results are elaborated in Section 6. We perform method evaluation in Section 7. Finally, we present the discussion and conclusions, respectively, in Section 8 and Section 9. To simplify, the acronyms used in this work are presented in Table 1.

## 2. Sparsity Issue in Recommender Systems

A CF recommender system presents items to users based on other individuals' choices. It assumes that individuals who had comparable choices previously are more probably to share similar preferences later [33]. The core concept of the CF method is that it collects data on users' choices of several products or services by referring to users' ratings [34]. Users with the same rating patterns are regarded as similar users, in which the similarity value is calculated by a particular algorithm such as K-Nearest Neighbor (KNN). Although the CF approach has earned researchers' interest recently, the performance of this method depends basically on the feedback exhibited by users. Still, users may not be concerned with providing feedback based on several causes which might lead to the data sparsity issue [35]. Data sparsity is a basic obstacle of the CF method [36]. The number of users, products, and services in any recommender system might impact the number of rated items by individuals [35]. Efficient estimation of ratings referring to a limited number of examples is significant [37]. Still, without sufficient information, it is difficult for the CF approach to present efficient recommendations to users. A sparsity issue emerges because the user interacts only with a limited set of products in a specific domain. For example, the MovieLens datasets entail a rating matrix, in which users provide their ratings of movies. Still, only 10% of this matrix is filled by users [6], which indicates the data sparsity problem.

Many methods have tried to address the data sparsity of RSs, which can be classified based on several aspects. The first method tries to utilize additional data of users. Additional data like friends' connections and consumer-generated tags can also be integrated into RSs to identify similar neighbors [38]. To address the data sparsity in collaborative filtering recommender systems, Guo et al. [39] designed the TrustSVD model, in which implicit and explicit impacts of social trust were utilized to estimate items for real users. Their proposed model surpassed trust-based and rating-based approaches in the accuracy of prediction among users with various trust levels. On the other hand, Tang et al. [40]

TABLE 1: List of acronyms in alphabetical order.

Acronym	Term
AIC	Akaike information criterion
ANFIS	Adaptive neurofuzzy inference system
ANN	Artificial neural network
BPR	Bayesian personalized ranking
CF	Collaborative filtering
EM	Expectation-maximization
FN	Fuzzy number
FIS	Fuzzy inference system
GMM	Gaussian mixture model
HOSVD	Higher-order singular value decomposition
SVD	Singular value decomposition
KNN	K-nearest neighbor
MAE	Mean absolute error
MC-CF	Multicriteria collaborative filtering
MF-BPR	Multifedback Bayesian personalized ranking
MFs	Membership functions
MRMF	Multirelational matrix factorization
NN	Nearest neighbor
PCA	Principle component analysis
R2	Coefficient of determination
RMSE	Root mean squared error
RS	Recommender system
TRs	Travel recommender systems

presented the SoDimRec model, in which heterogeneous social links and poor dependencies among them are considered. The research model presented a good performance based on real-world data sets. In another study by Krohn-Grimberghe, Drumond, Freudenthaler, and Schmidt-Thieme [41], the Multirelational Matrix Factorization (MRMF) technique was utilized using a Bayesian personalized ranking (BPR) framework based on several social connections among users for recommending items to users. The authors indicate the effectiveness of the presented technique in addressing the problem of the lack of social relations among users based on the real dataset. Zhao et al. [42] focused on the utilization of social relations to present accurate ranking paradigms. The authors assumed that a product purchased by an active consumer is more preferred than the product purchased by a friend, which is accordingly more preferred than a product purchased by other users. They designed the social Bayesian personalized ranking model referring to that assumption and presented promising findings based on the evaluation of real-world datasets. To enhance the performance of RSs in a sparsity context, Feng et al. [43] presented a multifactor similarity metric that locates nonlinear and linear connections among users from excessive behavior. The disadvantage of these methods is that there is a need for built-in social connections which may not be always available [44].

The second method tries to integrate additional data of items (features and content). McAuley and Leskovec [45] present the hidden factors as a topics model to integrate product factors using review texts aiming to enhance the prediction of rating. The proposed method can address the issue of new items with limited ratings. Zhu et al. [46] tried to address the data sparsity of ratings by referring to several

extracted side information of items from the social media. They also converted user-item ratings into weighted topic-item ratings. Compared to other methods and based on the evaluation of real-world datasets, the outcomes indicated the efficiency of the presented method. He and McAuley [47] deployed deep networks to investigate the influence of extracted visual characteristics of item images. The deployed approach could enhance the performance of Top-N product suggestions. In a study by Vasile et al. [48], categorical side data of products was utilized in the recommender system. The metadata was incorporated to frame the embedding representation of products. One limitation of this method is overspecialization, which leads to presenting a limited range of products to users. The research findings indicated that new product representations improved the performance of RSs based on a music dataset.

The third approach considered several interactions between users and products. Pan et al. [49] presented an adaptive Bayesian personalized ranking to address the problem of heterogeneous implicit feedback. The authors assumed that products with browse actions will be more favored than products without any. The outcomes of the study affirmed that the deployed method is capable of leveraging uncertainty in examination records efficiently compared to other several ranking-oriented assessment measures. Qiu et al. [50] differentiated consumers' preferences based on three classes of products: the purchased products, products with auxiliary behaviors, and products without any behavior. The authors developed a trinity preference-based BPR model aiming to improve the outcomes of RSs. Loni et al. [51] proposed the MF-BPR approach. They designed a nonlinear sampling technique for the standard BPR method, in which the sampling probability depends on the degree of positive feedback an individual may have on a product.

To get an insight into the current research regarding the sparsity problem in the recommender system, a visualization of the keyword cooccurrence network is generated through a bibliometric analysis. The main outcome of the bibliometric analysis that we performed is a cooccurrence-keywords map of the research topic. We used the following keywords to retrieve the related terms from the Scopus database: (“sparse data” OR “sparsity”) AND (“recommender system” OR “recommendation system” OR “recommender agent” OR “recommendation agent” OR “recommender engine” OR “recommendation engine”).

In Figure 1, the distance among the elements is established by calculating how many studies in which both items (keywords) occur. A huge number of cooccurrences is indicated by a short path among the represented items. That distance is indicated in the cooccurrence image which is utilized to present the segments of the keywords. Besides, bigger frames indicate more occurrences in the studies; hence, as the figure presents, “recommender systems,” “data sparsity,” and “collaborative filtering” are the most frequent keywords in the selected studies. The diagram entails four segments: segment 1 (14 keywords), segment 2 (12 keywords), segment 3 (10 keywords), and segment 4 (9 keywords). In segment 1 (red color), the core keywords are

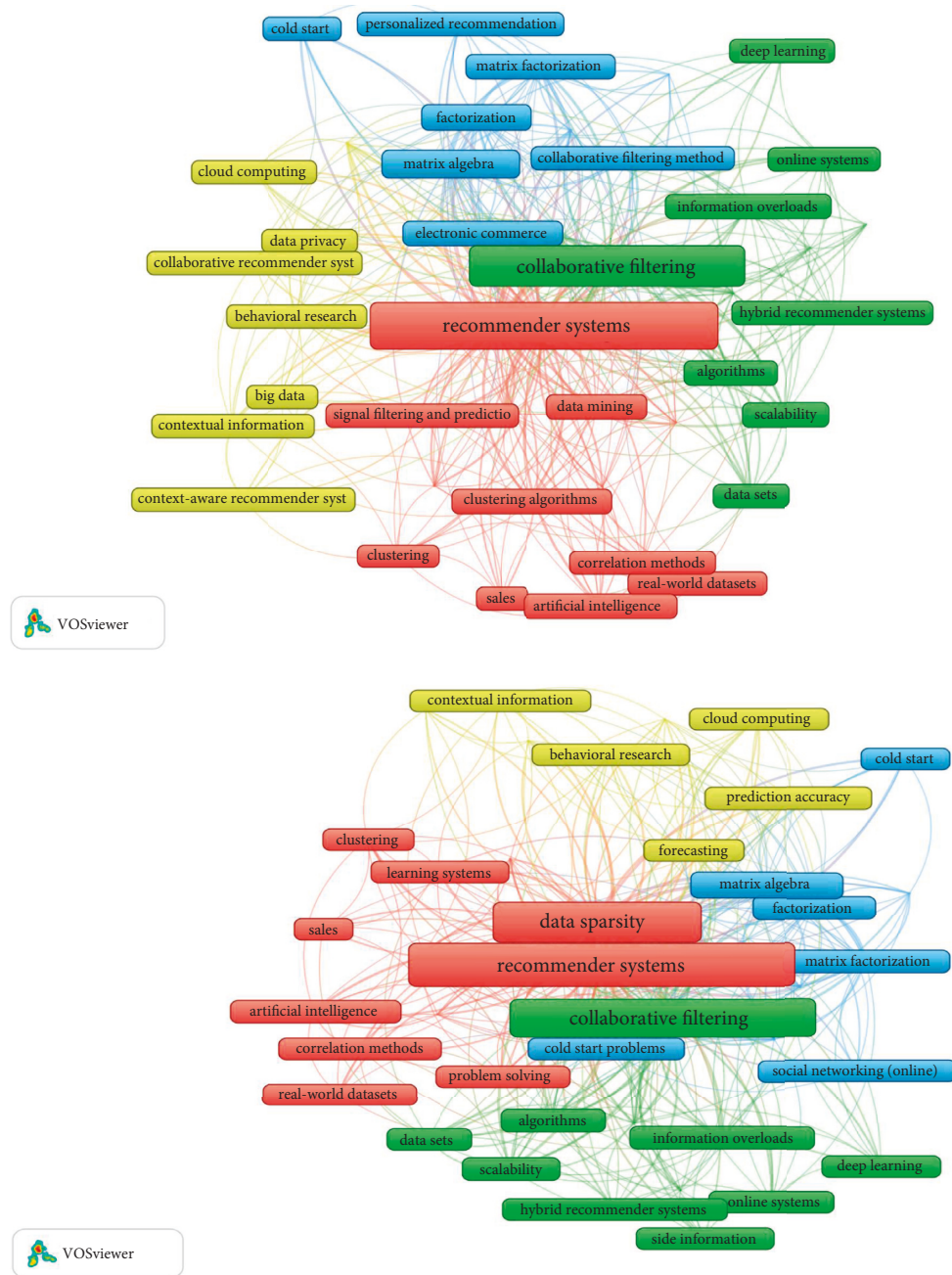


FIGURE 1: Cooccurrence-keywords visualization (based on 90-degree rotation of the diagram).

“recommender systems” and “data sparsity.” In segment 2 (green color), the core keyword is “collaborative filtering.” In segment 3 (blue color), the core keywords are “matrix algebra” and “factorization.” In segment 4 (yellow color), the core keyword is “prediction accuracy” and “context-aware recommender system.”

### 3. Related Work on Recommendation Agents in Tourism and Hospitality

Tourism is the movement of people from one geographical place (which is usually their place of residence) to another place to achieve various individual, business, and leisure

goals [52]. The most frequently used description of the tourism system is presented by Leiper [53] who indicated that the tourism system is a three-part system that entails production, transportation, and tourism destinations, in which these parts are placed within economic, social, and environmental contexts. Based on several business and marketing activities, tourism has a vital part in the economic, social, and cultural advancement of most regions [54]. An important feature of tourism activities is their close link with tourists’ preferences and interests [55, 56].

The fast advancement of social platforms allows online commerce to change from a product-based platform to a social-based one [57–59]. Thus, online business has

encountered an emerging formation that adopts Web 2.0 characteristics to allow consumers to be engaged and motivates them to communicate [60], which accordingly can add more financial value for businesses. Hence, the fast advancements in the Internet, communication technologies, and social media have had an intense impact on the tourism business [61]. Nowadays, tourists tend to utilize smart devices to help them in browsing and finding places that suit their aims [62]. In this social-based environment, tourists search for other tourists who shared their experiences and knowledge to increase their awareness of the presented services and, finally, reach the right choices. People tend to understand the opinions of other individuals and make choices under the impact of social ties [63, 64].

RSs deployed in tourism applications are widely recognized as destination recommendation systems or Travel Recommender Systems (TRSs) [65]. RSs in the tourism field are beneficial tools for consumers and travel agents [66]. Trying to imitate the interaction with a real tourism agent, several tourism providers have integrated RSs into their web portals. Using TRSs, tourists can simply reach related information about the locations they require, thus, leading to less time for booking decisions with more tailored suggestions that meet their preferences. TRS is an intelligent system that returns tourism services by presenting guidelines and suggestions to tourists. TRSs can be categorized as either web-based TRSs or mobile TRSs.

#### 4. Methodology

This paper develops a new method for the proposed recommender system through clustering, dimensionality reduction, and the neurofuzzy system (see Figure 2). In the first step of the method, we perform data segmentation using EM. There are many types of clustering techniques for data segmentation. It has been demonstrated that EM clustering techniques are robust when the dataset is sparse [67]. Through segmentation, similar users can be better detected with lower computation time. This also will help the ANFIS technique to better construct the prediction models when the data is large. In addition, in this stage, EM will be applied for the different number of clusters and the best cluster size will be selected for further analysis in the next step which is dimensionality reduction through HOSVD. HOSVD aims to reduce the dimensions of the data for similarity calculations between users and items in lower dimensions. In the next step, we use a neurofuzzy approach, ANFIS, to predict users' overall ratings according to the predicted criteria ratings. Accordingly, a dense tensor of the data will be used in the next stage of the method for the recommendation procedures. The method predicts the hotels' ratings for the users and recommends them according to the developed fuzzy-based algorithms. The results are finally evaluated using several metrics, Precision, F1-measure, RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error). The techniques

which have been used in the proposed recommendation method are introduced in the following sections.

*4.1. Prediction and Recommendation Procedure.* In this section, we first present core concepts of fuzzy sets [68], which are used to complete the prediction and recommendation procedure in the proposed method.

*Definition 1.* For a fuzzy number (FN)  $\tilde{a}$ , the  $\lambda$ -cut for  $\mu_{\tilde{A}}^{-}(x)$ ,  $x \in [0,1]$ , in fuzzy set  $\tilde{A}$   $\tilde{a}_{\lambda} = [a_{\lambda}^{-}, a_{\lambda}^{+}]$  ( $a_{\lambda}^{+}$  indicates the upper bounds and  $a_{\lambda}^{-}$  indicates the lower bounds of the closed interval) is defined as

$$\tilde{a}_{\lambda} = \{x, \mu_{\tilde{a}}^{-}(x) \geq \lambda, x \in R\}. \quad (1)$$

*Definition 2.* We define the membership function (MF)  $\mu_{\tilde{a}}^{-}(x)$  for a triangular FN =  $\tilde{a}$  through triplet  $(a_0^{-}, a, a_0^{+})$  as

$$\mu_{\tilde{a}}^{-}(x) = \begin{cases} 0, & X < a_0^{-}, \\ \frac{x - a_0^{-}}{a - a_0^{-}}, & a_0^{-} \leq X \leq a, \\ \frac{a_0^{+} - x}{a_0^{+} - a}, & a \leq X \leq a_0^{+}, \\ 0, & a_0^{+} < X. \end{cases} \quad (2)$$

*Definition 3.* For any  $\tilde{a}, \tilde{b} \in F_{+}^{*}(R)$ , the group of all finite positive FN on  $R$ , and  $0 < \delta \in R$  we have

$$\begin{aligned} \tilde{a} + \tilde{b} &= \bigcup_{\lambda \in [0,1]} \lambda [a_{\lambda}^{-} + b_{\lambda}^{-}, a_{\lambda}^{+} + b_{\lambda}^{+}], \\ \delta \tilde{a} &= \bigcup_{\lambda \in [0,1]} \lambda [\delta a_{\lambda}^{-}, \delta a_{\lambda}^{+}], \\ \tilde{a} \times \tilde{b} &= \bigcup_{\lambda \in [0,1]} \lambda [a_{\lambda}^{-} \times b_{\lambda}^{-}, a_{\lambda}^{+} \times b_{\lambda}^{+}]. \end{aligned} \quad (3)$$

*Definition 4.* Suppose  $\tilde{a}$  and  $\tilde{b}$  are two FNs. Then  $\tilde{a} = \tilde{b}$  if  $a_{\lambda}^{-} = b_{\lambda}^{-}$  and  $a_{\lambda}^{+} = b_{\lambda}^{+}$  for any  $\lambda \in [0,1]$ .

Learning the prediction functions for the items as well as users has been taken into account in the present work and following prior works on multicriteria collaborative filtering (MC-CF). A weighted approach has been used in every cluster to obtain a combination of the prediction functions. According to previous studies [69, 70], we utilize equation (4) as a common weighting scheme, which is represented as

$$\hat{r}_{u,x} = w_u * \hat{r}_{u,x}^{\text{user}} + w_x * \hat{r}_{u,x}^{\text{item}}. \quad (4)$$

In equation (4),  $w_u$  and  $w_x$  are the weight of  $\hat{r}_{u,x}^{\text{user}}$  and  $\hat{r}_{u,x}^{\text{item}}$ , respectively. According to the definitions of fuzzy set, we extend the above prediction method for fuzzy-based prediction.

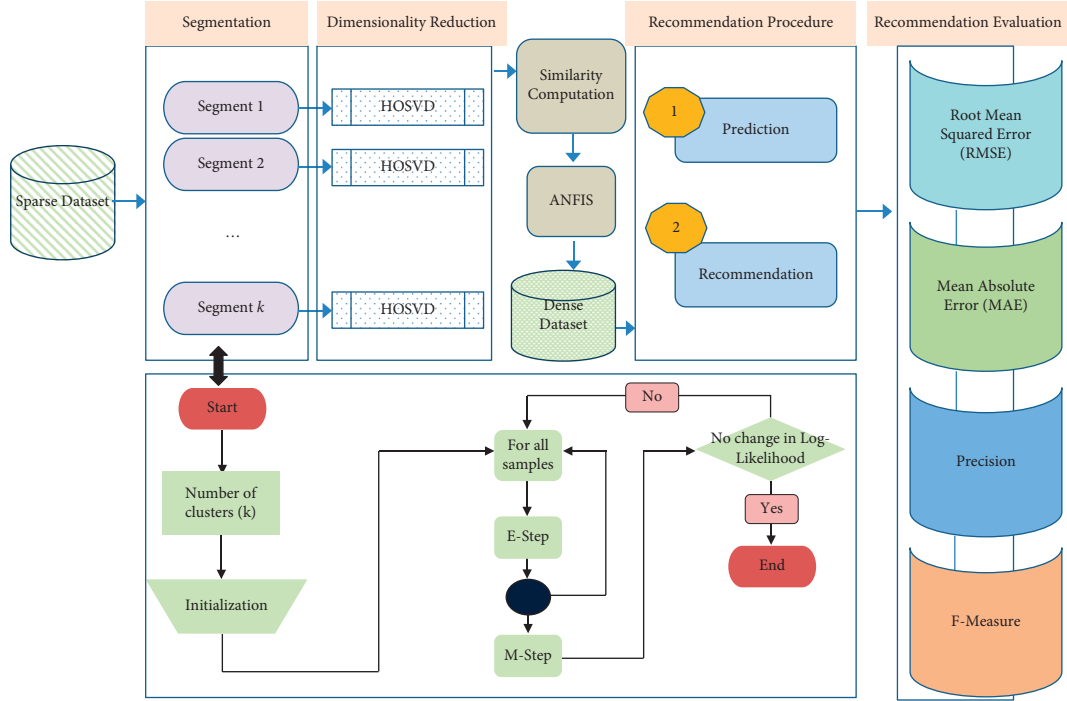


FIGURE 2: Research method.

$$\hat{r}_{u,x} = w_u * \hat{r}_{u,x}^{\text{user}} + w_x * \hat{r}_{u,x}^{\text{item}} = \bigcup_{\alpha \in [0,1]} \alpha [w_u \times^{\text{user}} \hat{r}_{u,x}^- + w_x \times^{\text{item}} \hat{r}_{u,x}^-, w_u \times^{\text{user}} \hat{r}_{u,x}^+ + w_x \times^{\text{item}} \hat{r}_{u,x}^+]. \quad (5)$$

*Definition 5.* As defined by Adomavicius and Kwon [71], a multicriteria recommender system incorporates preferences into multiple criteria or dimensions which provide more information about the items and users. For more literature on multicriteria recommender systems see the previous studies by Adomavicius and Kwon [70, 71].

Sparsity has been a major disadvantage of many collaborative filtering RSs which can significantly impact the accuracy of items' recommendations. CF algorithm generates inefficient recommendations when there are a lower number of ratings. Enough amounts of rating data are required by CF recommendation algorithms. Clustering as well as fuzzy rule-based methods has been used in the present work to deal with this problem. Clustering techniques are aimed to improve the efficiency of

recommendation agents [70, 72]. The fuzzy logic approach is demonstrated to be effective in solving the sparsity issue in CF-based recommender systems [23, 32, 73]. Moreover, the technique proposed by Adomavicius and Kwon [71] was used in the present work to establish similarities of users (see equation (6)), after which it was applied as a fuzzy-based similarity computation technique (see equation (7)). The average similarity of the two users is achieved by the deployment of the suggested method according to

$$\text{sim}_{\text{avg}}(u, u') = \frac{1}{k+1} \sum_{c=0}^k \text{sim}_c(u, u'). \quad (6)$$

According to the definition based on the fuzzy set, we define  $\text{sim}_c(u, u')$  as [74]

$$\text{sim}_c(u, u') = \frac{\sum_{x \in I_{u,u'}} \int_0^1 (1/2) [(r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^-)(r_{x,u_\alpha}'^- - \bar{r}_{x,u_\alpha}'^-) + (r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^-)(r_{x,u_\alpha}'^+ - \bar{r}_{x,u_\alpha}'^+)] d\alpha}{\sqrt{\sum_{x \in I_{u,u'}} \left( \int_0^1 (1/2) [(r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^-)(r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^+)] d\alpha \right)^2} \times \sqrt{\sum_{x \in I_{u,u'}} \left( \int_0^1 (1/2) [(r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^+)(r_{x,u_\alpha}^- - \bar{r}_{x,u_\alpha}^+)] d\alpha \right)^2}} \quad (7)$$

The recommending system initially discovers the active or target users along with the active or target hotels in the online recommendation phase. The tasks of rating predictions as well as recommendations are carried out in the next

stage. The first task includes every algorithm for the prediction of the ratings associated with a list of hotels according to the determined priorities of a specific active user. In the second phase, the system includes the ranking of

a list of items that have not been rated for active users after which the Top-N recommendations, including the first  $N$  hotels in the list of recommendations, are provided.

**4.2. Data Segmentation.** As a probabilistic method, Gaussian Mixture Model (GMM) uses parametric Gaussian

distribution to illustrate each cluster of the data. Accordingly, a linear superposition of Gaussian components is used to model  $f(x)$  which is the distribution of the entire dataset (see equation (8)).

$$f(x) = \sum_{j=1}^J p_j f(x|\mu_j, \Sigma_j) = \sum_{j=1}^J p_j \frac{\exp\left(-1/2(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)\right)}{(2\pi)^{d/2} |\Sigma_j|^{1/2}}. \quad (8)$$

$\sum_{j=1}^J p_j = 1$  and  $\int_x f(x|\Sigma_j, \mu_j) dx = 1$ . Referring to the GMM, there are  $J$  distributions that allow the establishment of  $f(x)$ , and maximum-likelihood is utilized for the estimation of the unknown parameters,  $\Theta = \{\mu_j, \Sigma_j, p_j\}_{j=1}^J$  through the maximization of the log-likelihood function referring to the group of the available training samples  $\{x_k\}_{k=1}^N$  as shown in the following equation:

$$L = \sum_{k=1}^N \log \left( \sum_{j=1}^J p_j f(x_k|\mu_j, \Sigma_j) \right). \quad (9)$$

It has been shown that the EM algorithm is effective in data clustering in which the available dataset is incomplete.

The GMM parameters in the EM algorithm are computed as follows:

- (1) Initialization: The initial estimates  $p_j(0)$ ,  $\mu_j(0)$ ,  $\Sigma_j(0)$ ,  $j = 1, 2, \dots, J$  are selected and then the preliminary log-likelihood is computed as follows:

$$L(0) = \sum_{k=1}^N \log \left( \sum_{j=1}^J p_j(0) f(x_k|\mu_j(0), \Sigma_j(0)) \right). \quad (10)$$

- (2) Expectation Step (E-Step): In this step, we compute

$$p(j|x_k, \Theta(t)) = \frac{p_j(t) f(x_k|\mu_j(t), \Sigma_j(t))}{\sum_{l=1}^J p_l(t) f(x_k|\mu_l(t), \Sigma_l(t))}, \quad j = 1, 2, \dots, j, k = 1, 2, \dots, N, \quad (11)$$

$$p_j(t+1) = \frac{1}{N} \sum_{k=1}^N p(j|x_k, \Theta(t)).$$

- (3) Maximization Step (M-Step): In this step, the new estimate is computed as

$$\begin{aligned} \mu_j(t+1) &= \frac{\sum_{k=1}^N x_k p(j|x_k, \Theta(t))}{\sum_{k=1}^N p(j|x_k, \Theta(t))}, \\ \Sigma_j(t+1) &= \frac{\sum_{k=1}^N p(j|x_k, \Theta(t)) N(x_k - \mu_j(t+1))(x_k - \mu_j(t+1))^T}{\sum_{k=1}^N p(j|x_k, \Theta(t))}. \end{aligned} \quad (12)$$

- (4) Convergence Check: In this step, the new log-likelihood is computed as

$$L(t+1) = \sum_{k=1}^N \log \left( \sum_{j=1}^J p_j(t+1) f(x_k|\mu_j(t+1), \Sigma_j(t+1)) \right). \quad (13)$$

- (5) Return to the second step if  $|L(t+1) - L(t)| > \delta$  for a predetermined threshold  $\delta$ ; else end the algorithm.

**4.3. HOSVD.** HOSVD is an extension of the classical Singular Value Decomposition (SVD) to tensors. Multilinear rank is used in the HOSVD. In this study, HOSVD aims to

discover the latent relationships among the entities in a 3-order tensor  $A \in R^{|U||I||C|}$ . The tensor stores the data in 3 dimensions including hotels, hotels' criteria, and users (see Figure 3). The decomposition of the tensor is performed by unfolding the tensor to have 2D matrices  $A_1$ ,  $A_2$ , and  $A_3$ , as

$$\begin{aligned} A_1 &= U^{(1)} \cdot S_1 \cdot V_1^T, \\ A_2 &= U^{(2)} \cdot S_2 \cdot V_2^T, \\ A_3 &= U^{(3)} \cdot S_3 \cdot V_3^T. \end{aligned} \quad (14)$$

The core tensor is obtained by the following equation and using left singular vectors of the  $A_1$ ,  $A_2$ , and  $A_3$ :

$$S = A \times_1 U_1^{(1)T} \times_2 U_2^{(2)T} \times_3 U_3^{(3)T}. \quad (15)$$

Accordingly, we can obtain the approximation of the tensor as

$$\tilde{A} = S \times_1 U_1^{(1)} \times_2 U_2^{(2)} \times_3 U_3^{(3)}. \quad (16)$$

To find similar users in each cluster to perform the neighborhood formation, we used the cosine similarity formula for two vectors  $X$  and  $Y$  as provided in the following equation:

$$\text{Similarity}_c(A, B) = \frac{X \cdot Y}{\|X\| \cdot \|Y\|} = \frac{\sum_{i=1}^n X_i \times Y_i}{\sqrt{\sum_{i=1}^n X_i^2} \times \sqrt{\sum_{i=1}^n Y_i^2}}, \quad (17)$$

where  $X$  and  $Y$ , respectively, represent the Euclidean norm of vectors  $X$  and  $Y$ .

**4.4. ANFIS.** ANFIS model was presented by [75] by integrating Artificial Neural Network (ANN) and Fuzzy Inference System (FIS). ANFIS overcomes the shortcomings of ANN and FIS, like overfitting and sensitivity to the definition of membership functions, to present a better outcome concerning the prediction issues. The most general technique for ANFIS training is the Sugeno-type FIS, in which a robust learning algorithm is utilized to choose the parameters of the model. ANFIS is structured from five layers (see Figure 4), in which each layer entails several nodes as in ANN. In ANFIS, several steps are involved as input data fuzzification, constructing the fuzzy database, constructing the fuzzy rule base, development of the decision, and presenting the data defuzzification.

## 5. Data Collection and Analysis

The data collection procedure in this research was performed on TripAdvisor. Numerical ratings of hotels from travelers in Malaysia were considered. The frequency of ratings in different years is presented in Table 2. The numerical ratings indicate travelers' selections considering various hotels' characteristics. TripAdvisor allows travelers to evaluate each of six significant folds for each hotel by the travelers. The main folds are "Service," "Cleanliness," "Value," "Location," "Rooms," and "Sleep Quality."

Correlation among the criteria is presented in Table 3. Travelers can also indicate their degree of satisfaction with

the quality of the presented service based on the six main characteristics. The ratings are gathered from ten hotels in Malaysia on the TripAdvisor website from 2015 through 2021. A predesigned web-based crawler was used to gather the data, in which 28173 ratings were crawled. The gathered data was cleaned by removing ratings that only entail overall ratings leading to 22506 ratings. Users' rating information for overall and criteria ratings is presented in Table 4.

## 6. Result and Discussion

In this study, EM was chosen as a clustering approach for data segmentation. Choosing the appropriate number of segments is a vital issue in any clustering technique. The right number of segments should be chosen carefully to present segments with the best quality. In the EM method, the maximizing of likelihood is vital. This can be achieved using Akaike Information Criterion or AIC, as a model selection approach. Hence, we applied AIC to determine the optimal number of segments in the Expectation-Maximization algorithm. Besides, we used the 10-fold cross-validation approach in the Expectation-Maximization clustering process to get an unbiased outcome. The results of the segmentation by EM for TripAdvisor are presented in Table 5 with the best value AIC (478069.6732) for 8 segments of the EM. 1-way ANOVA results for clusters versus input attributes are also presented in Table 5. The results of the test indicated the statistical significance of differences in the means among the segments. The number of ratings for 8 clusters is 2200, 2960, 3213, 2827, 2963, 1848, 3205, and 3290, respectively, in segment 1 to segment 8. The clustering quality criterion and cluster centroids are presented in Table 6. According to the table, eight cluster centroids are provided. The lowest values for cluster centroids for the majority of the criteria belong to segment 6. It is found that the highest values belong to segment 1 for the majority of the criteria. In Table 7, clustering quality criteria based on overall ratings are presented.

MATLAB R2020 software was utilized in this research. We performed HOSVD in each cluster to solve the sparsity issue. By the use of the HOSVD technique, the neighborhood was constructed to find similar users in each cluster for unknown ratings. Then we applied ANFIS in each cluster to construct the prediction models. To build the models in the MATLAB Fuzzy Logic toolbox, there are two widely used approaches: backpropagation and hybrid. As the number of inputs increases in the ANFIS model, the MF shows exponential growth, thereby causing a high computation to develop the models. Each input has three MFs assigned to it and "linear" was chosen as the "MF Type" for the output. The ANFIS structure becomes more complex with the increasing number of input MFs. Hence, converging to the target error needs more iteration and the training process requires more time. The outcomes of the defuzzification stage are then utilized for the overall rating predictions. Following the application of Principle Component Analysis (PCA) on clusters, ANFIS models were deployed to locate the relative significance of criteria and to estimate the overall ratings according to the performance criteria Rooms, Value,



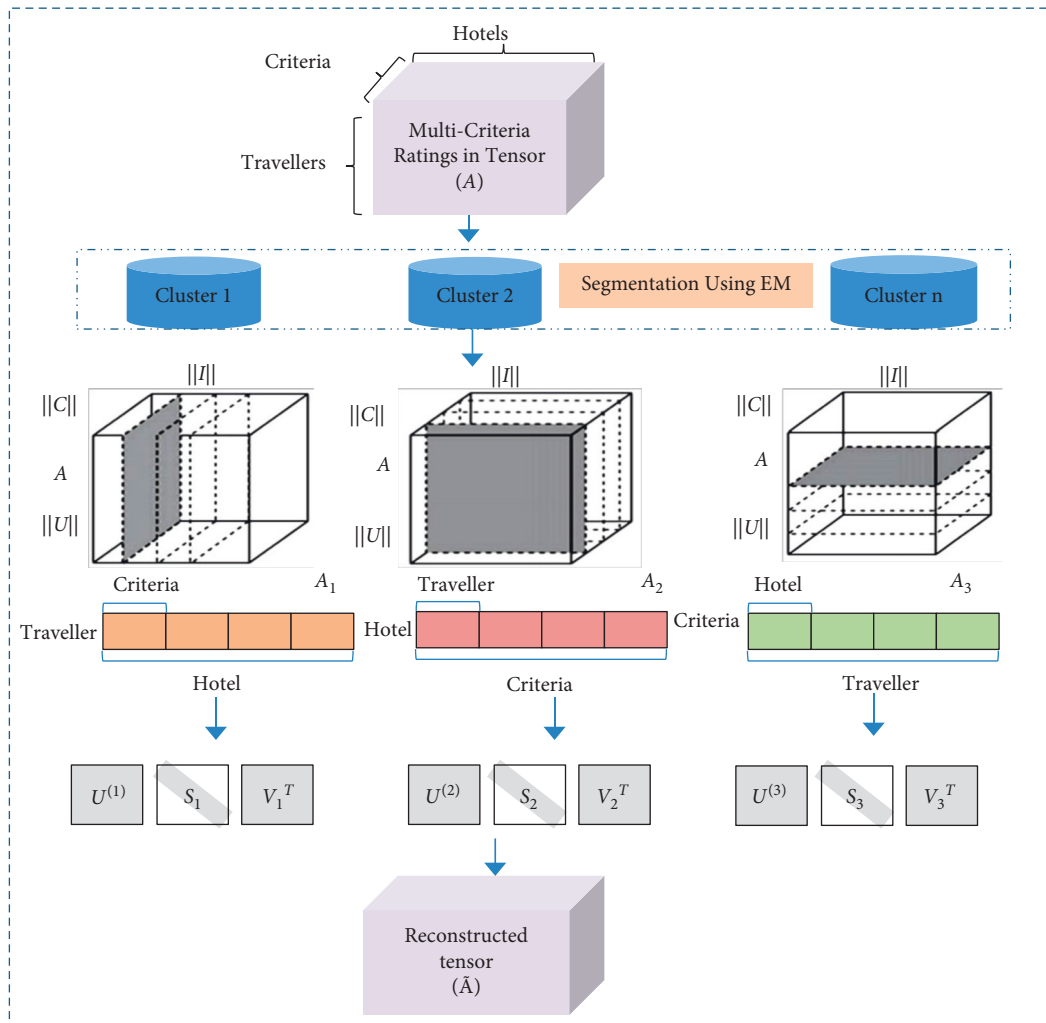


FIGURE 3: The process of HOSVD for tensor factorization.

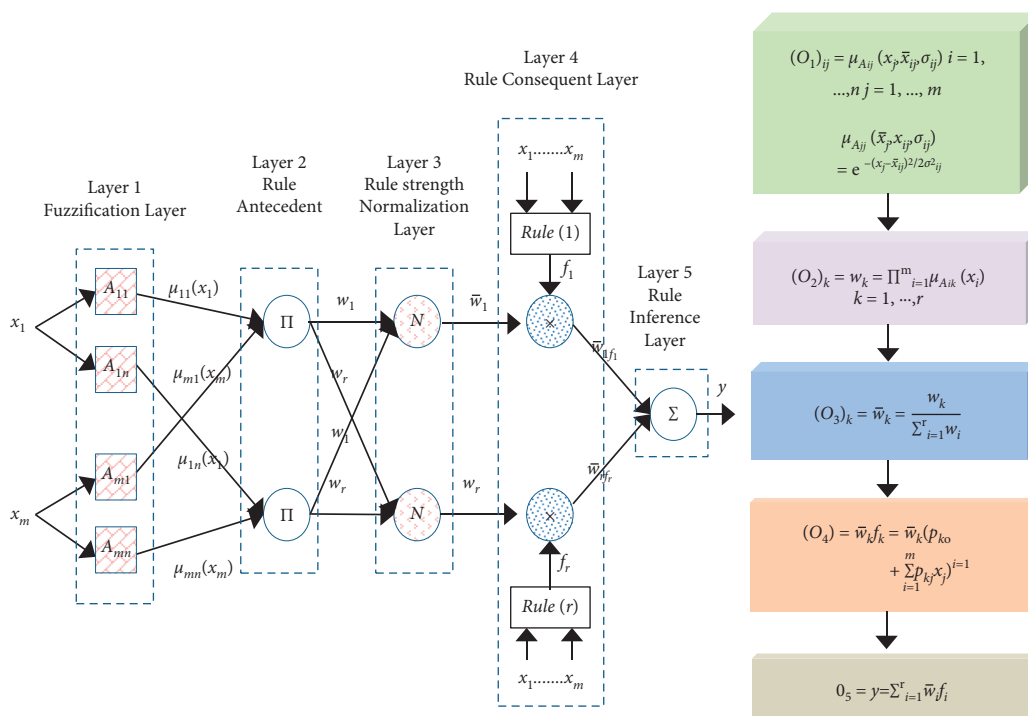


FIGURE 4: ANFIS architecture with its procedure.

TABLE 2: Frequency of ratings in different years.

Year	Overall rating					Total	Percent
	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)		
2015	40	312	1594	1614	379	3939	17.5
2016	31	306	1531	1580	340	3788	16.8
2017	53	352	1815	1746	432	4398	19.5
2018	40	241	1403	1405	336	3425	15.2
2019	100	460	1816	1729	387	4492	20.0
2020	40	123	654	679	167	1663	7.4
2021	7	66	318	333	77	801	3.6
Total	311	1860	9131	9086	2118	22506	100.0

TABLE 3: Correlation among the criteria.

	Correlations					
	Rooms	Value	Location	Service	Cleanliness	Sleep quality
Rooms	1	0.059**	0.048**	0.047**	0.054**	0.043**
Value	0.059**	1	0.039**	0.044**	0.053**	0.054**
Location	0.048**	0.039**	1	0.057**	0.054**	0.051**
Service	0.047**	0.044**	0.057**	1	0.051**	0.046**
Cleanliness	0.054**	0.053**	0.054**	0.051**	1	0.046**
Sleep quality	0.043**	0.054**	0.051**	0.046**	0.046**	1

\*\*Correlation is significant at the 0.01 level (2-tailed).

TABLE 4: Users' rating information for overall and criteria ratings.

		Overall rating				
		Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Rooms	Very low (1)	294	763	2231	1188	97
	Low (2)	12	478	2202	1534	172
	Moderate (3)	4	326	1860	1873	290
	High (4)	1	190	1547	2216	636
	Very high (5)	0	103	1291	2275	923
Value	Very low (1)	285	754	2280	1269	103
	Low (2)	15	485	2178	1572	178
	Moderate (3)	9	322	1866	1826	310
	High (4)	2	192	1567	2176	618
	Very high (5)	0	107	1240	2243	909
Location	Very low (1)	290	679	2232	1182	87
	Low (2)	13	537	2143	1541	164
	Moderate (3)	5	328	1852	1814	278
	High (4)	3	221	1563	2280	603
	Very high (5)	0	95	1341	2269	986
Service	Very low (1)	295	720	2303	1280	96
	Low (2)	14	517	2109	1518	183
	Moderate (3)	2	310	1845	1916	294
	High (4)	0	192	1605	2188	629
	Very high (5)	0	121	1269	2184	916
Cleanliness	Very low (1)	293	751	2258	1232	99
	Low (2)	13	495	2119	1537	173
	Moderate (3)	3	339	1941	1900	265
	High (4)	2	173	1556	2205	631
	Very high (5)	0	102	1257	2212	950
Sleep quality	Very low (1)	291	676	2245	1300	106
	Low (2)	9	525	2130	1487	173
	Moderate (3)	10	341	1937	1851	269
	High (4)	1	201	1519	2208	637
	Very high (5)	0	117	1300	2240	933

TABLE 5: 1-way ANOVA for clusters versus input attributes.

Attribute Y	Attribute X	Description				Statistical test		
		Value	Examples	Average	Std-de	Variance decomposition		
Rooms	EM cluster	Segment-1	2200	3.8573	1.0203	Source	Sum of square	d.f.
		Segment-2	2960	3.9125	1.0210	BSS	17876.3754	7
		Segment-3	3213	2.5584	1.3308	WSS	27769.2741	22498
		Segment-4	2827	4.2565	0.7451	TSS	45645.6495	22505
		Segment-5	2963	1.6369	0.7105	Significance level		
		Segment-6	1848	1.7413	0.7809	Statistics	Value	Proba
		Segment-7	3205	2.7735	1.3757	Fisher's F	2069.001535	≤0.001
		Segment-8	3290	3.1827	1.4007			
		All	22506	3.0102	1.4242			
Value	EM cluster	Segment-1	2200	4.0577	0.8930	Source	Sum of square	d.f.
		Segment-2	2960	1.9155	0.8963	BSS	24632.8757	7
		Segment-3	3213	3.7893	1.1622	WSS	21107.1895	22498
		Segment-4	2827	1.8599	0.8723	TSS	45740.0653	22505
		Segment-5	2963	1.9433	0.9192	Significance level		
		Segment-6	1848	1.9464	0.9347	Statistics	Value	Proba
		Segment-7	3205	3.5189	1.2520	Fisher's F	3750.857613	≤0.001
		Segment-8	3290	4.4371	0.6405			
		All	22506	2.9886	1.4256			
Location	EM cluster	Segment-1	2200	4.2050	0.7690	Source	Sum of square	d.f.
		Segment-2	2960	2.1064	1.0750	BSS	21800.7256	7
		Segment-3	3213	1.7077	0.7495	WSS	23888.6228	22498
		Segment-4	2827	3.4574	1.2876	TSS	45689.3484	22505
		Segment-5	2963	3.8157	1.1281	Significance level		
		Segment-6	1848	1.8874	0.9151	Statistics	Value	Proba
		Segment-7	3205	4.3339	0.7020	Fisher's F	2933.092161	≤0.001
		Segment-8	3290	2.6751	1.3261			
		All	22506	3.0317	1.4248			
Service	EM cluster	Segment-1	2200	4.1536	0.7375	Source	Sum of square	d.f.
		Segment-2	2960	2.7689	1.3658	BSS	16952.6087	7
		Segment-3	3213	3.0479	1.3784	WSS	28567.3897	22498
		Segment-4	2827	3.1772	1.3609	TSS	45519.9984	22505
		Segment-5	2963	4.0192	0.9300	Significance level		
		Segment-6	1848	1.8718	0.9482	Statistics	Value	Proba
		Segment-7	3205	1.4555	0.5539	Fisher's F	1907.268560	≤0.001
		Segment-8	3290	3.4596	1.2705			
		All	22506	3.0003	1.4222			
Cleanliness	EM cluster	Segment-1	2200	4.1536	0.7375	Source	Sum of square	d.f.
		Segment-2	2960	2.7689	1.3658	BSS	16952.6087	7
		Segment-3	3213	3.0479	1.3784	WSS	28567.3897	22498
		Segment-4	2827	3.1772	1.3609	TSS	45519.9984	22505
		Segment-5	2963	4.0192	0.9300	Significance level		
		Segment-6	1848	1.8718	0.9482	Statistics	Value	Proba
		Segment-7	3205	1.4555	0.5539	Fisher's F	1907.268560	≤0.001
		Segment-8	3290	3.4596	1.2705			
		All	22506	3.0003	1.4222			
Sleep quality	EM cluster	Segment-1	2200	3.8932	1.0205	Source	Sum of square	d.f.
		Segment-2	2960	4.2929	0.7456	BSS	14783.2298	7
		Segment-3	3213	2.5387	1.3247	WSS	30937.2330	22498
		Segment-4	2827	1.6314	0.6834	TSS	45720.4628	22505
		Segment-5	2963	3.2973	1.3514	Significance level		
		Segment-6	1848	2.0568	1.0512	Statistics	Value	Proba
		Segment-7	3205	3.2424	1.3756	Fisher's F	1535.796704	≤0.001
		Segment-8	3290	2.9483	1.4094			
		All	22506	3.0083	1.4253			

TABLE 6: Clustering quality criterion and cluster centroids.

Attribute	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8
Rooms	3.857273	3.912500	2.558357	4.256456	1.636855	1.741342	2.773479	3.182675
Value	4.057727	1.915541	3.789293	1.859922	1.943301	1.946429	3.518877	4.437082
Location	4.205000	2.106419	1.707750	3.457375	3.815727	1.887446	4.333853	2.675076
Service	4.052727	2.989189	4.319950	2.980545	2.995275	1.817641	3.101092	1.562310
Cleanliness	4.153636	2.768919	3.047930	3.177220	4.019237	1.871753	1.455538	3.459574
Sleep quality	3.893182	4.292905	2.538749	1.631411	3.297334	2.056818	3.242434	2.948328

Log-likelihood = -238811.8366, AIC = 478069.6732, and BIC = 479858.4760.

TABLE 7: Clustering quality criterion and cluster centroids based on overall ratings.

Criteria and overall ratings	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8	
Rooms	Very low (1)	45	47	884	0	1465	850	758	524
	Low (2)	202	251	850	2	1125	640	721	607
	Moderate (3)	448	648	645	509	357	344	697	705
	High (4)	832	982	469	1078	16	14	547	652
	Very high (5)	673	1032	365	1238	0	0	482	802
Value	Very low (1)	14	1174	158	1190	1158	758	239	0
	Low (2)	111	1005	325	954	992	532	509	0
	Moderate (3)	400	645	656	572	641	457	692	270
	High (4)	884	129	971	111	167	101	880	1312
	Very high (5)	791	7	1103	0	5	0	885	1708
Location	Very low (1)	0	1042	1473	271	112	789	0	783
	Low (2)	25	993	1242	417	305	578	1	837
	Moderate (3)	396	588	462	649	621	381	430	750
	High (4)	882	242	36	728	904	100	1272	506
	Very high (5)	897	95	0	762	1021	0	1502	414
Service	Very low (1)	0	541	0	522	569	827	539	1696
	Low (2)	65	593	0	582	590	572	601	1338
	Moderate (3)	501	700	476	659	637	408	730	256
	High (4)	887	609	1233	557	620	41	667	0
	Very high (5)	747	517	1504	507	547	0	668	0
Cleanliness	Very low (1)	0	678	557	417	0	838	1839	304
	Low (2)	0	710	659	528	173	532	1272	463
	Moderate (3)	455	627	718	651	734	358	94	811
	High (4)	952	508	631	599	919	117	0	841
	Very high (5)	793	437	648	632	1137	3	0	871
Sleep quality	Very low (1)	41	0	915	1373	402	715	479	693
	Low (2)	207	0	816	1123	461	533	532	652
	Moderate (3)	408	516	658	331	682	416	695	702
	High (4)	834	1061	484	0	690	148	731	618
	Very high (5)	710	1383	340	0	728	36	768	625
Overall rating	Very low (1)	0	0	0	1	0	309	0	1
	Low (2)	0	148	217	238	194	732	168	163
	Moderate (3)	77	1336	1431	1415	1403	765	1299	1405
	High (4)	1004	1313	1396	1065	1246	42	1509	1511
	Very high (5)	1119	163	169	108	120	0	229	210

Location, Service, Cleanliness, and Sleep Quality. We used the hybrid and backpropagation learning approaches in the ANFIS technique. Totally, eight ANFIS prediction models were created upon the input and output data for the segments. Principal Components (PCs) were chosen as inputs of ANFIS models, in the fuzzification stages. Hence, suitable MFs were determined for all PCs. We selected triangular MFs and

created the ANFIS models based on this selection. This type of MF was shown to be effective in developing recommendation systems [32, 76]. Other types of MFs such as trapezoidal and Gaussian MFs can also be used in developing the prediction models.

Different models were constructed using MFs, i.e., triangular MFs, which represent triangular MF (see Figure 5). This MF is described by the following equation:

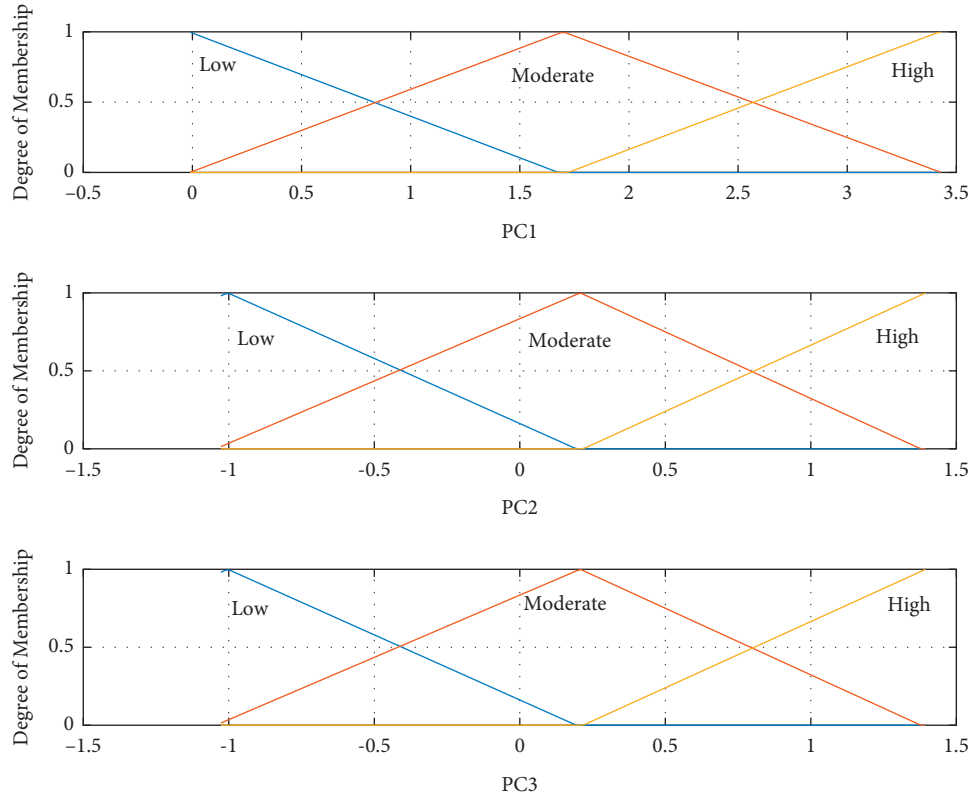


FIGURE 5: Membership functions in ANFIS models.

$$\text{trimf} = f(x; a, b, c) = \left\{ \begin{array}{ll} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ \frac{c-x}{c-b}, & b \leq x \leq c, \\ 0, & c \leq x. \end{array} \right. \quad (18)$$

## 7. Method Evaluation and Comparisons

We deployed ANFIS to model item prediction models from TripAdvisor ratings. The deployed method has several advantages over previous methods. First, we used ANFIS for learning the prediction models from the training datasets. We decided to follow this technique because ANFIS can predict overall ratings with higher accuracy. In addition, ANFIS is flexible in inducing fuzzy rules from the ratings in each segment. Second, the ANFIS has a flexible architecture which makes it appropriate in presenting stipulated input-output pairs by utilizing a group of induced fuzzy IF-THEN rules with suitable and various MFs.

In this study,  $R^2$  was deployed to identify the correlation factor, which is measured by equation (20).  $R^2$  is the proportion of variance in the observed data that is described by the model.  $R^2$  values fall in the interval of 0–1, as higher values indicate more capability of explaining the variance.

Another assessment, i.e., RMSE, was also utilized to measure the error of the training, validation, testing, and all datasets. The following formulas were used in the evaluation process:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i,j=1}^n (y_j - y_i)^2}, \quad (19)$$

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = \frac{\sum_j (y_j - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}, \quad (20)$$

where  $y_j$  indicates the predicted labels, and  $y_i$  indicates the true labels.  $SS_{\text{reg}}$  is the regression sum of squares (i.e., explained sum of squares), and  $SS_{\text{tot}}$  is the total sum of squares, which is proportional to the variance of the data.

Figures 6(a) and 6(b) indicate the RMSE of training for various epochs for various segments of EM for both hybrid and backpropagation methods. As we have used EM for the data segmentation, a total of 8 prediction models were constructed according to the number of segments generated by EM. For the error of prediction in the EM clusters, after 200 epochs, the averages RMSE, MAE, and  $R^2$  were measured. The results are presented in Table 8. It is found that the hybrid approach in ANFIS provides better RMSE and  $R^2$  values in training, testing, checking, and also all data.

To evaluate the deployed techniques compared with previously used approaches in MC-CF, we utilized the precision and recall metrics, which are broadly utilized to evaluate the quality of collaborative filtering. F-measure,

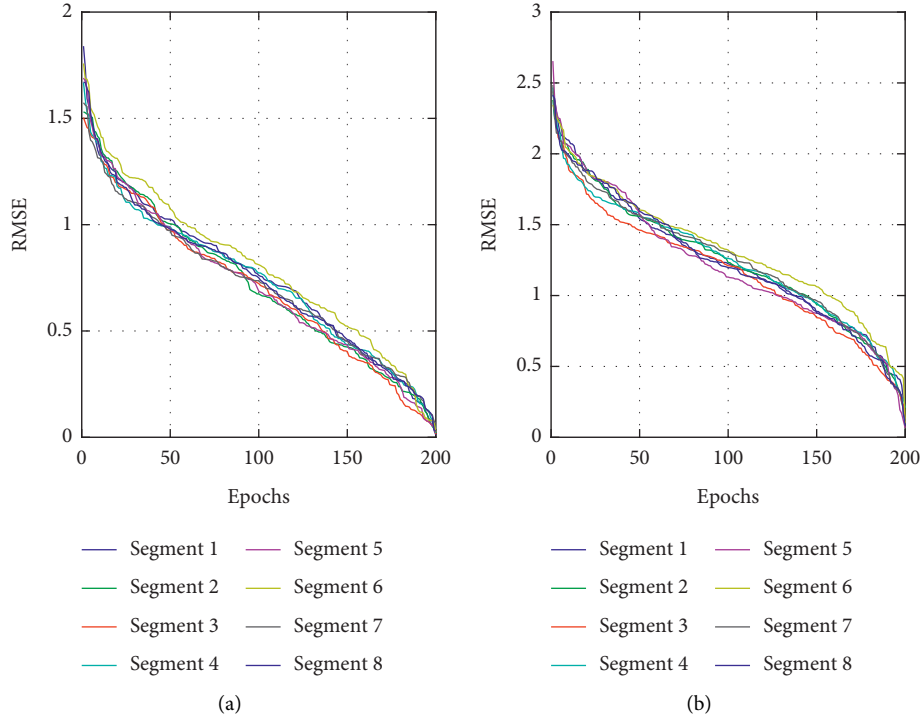


FIGURE 6: RMSE in eight clusters of EM in 200 epochs. (a) Hybrid and (b) backpropagation.

TABLE 8: Statistical values for prediction of overall ratings.

Approach	Dataset	RMSE	$R^2$
Hybrid	Training dataset	0.0280	0.9732
	Testing dataset	0.0421	0.9514
	Checking dataset	0.0782	0.9386
	All dataset	0.0274	0.9767
Backpropagation	Training dataset	0.1629	0.8721
	Testing dataset	0.1767	0.8663
	Checking dataset	0.1942	0.8482
	All dataset	0.1689	0.8771

which is defined as the harmonic mean of the recall and precision, is broadly utilized to assess the quality of RSs. The recall and precision measures are provided in equation (21) and equation (22). In equation (23), we present F-measure [74].

$$\text{Precision} = \frac{TR}{TR + FR}, \quad (21)$$

$$\text{Recall} = \frac{TR}{TR + FN}, \quad (22)$$

$$\begin{aligned}
 F &= \frac{1}{\beta^2/1 + \beta^2 \times 1/\text{Precision} + 1/1 + \beta^2 \times 1/\text{Recall}} \\
 &= \frac{(1 + \beta^2) \cdot p \times r}{\beta^2 \cdot \text{Recall} + \text{Precision}}.
 \end{aligned} \quad (23)$$

True relevant or TR represents the number of truly related predictions (how many suggested items as relevant

that are truly relevant). Also, false relevant or FR represents the number of suggested items that are supposed to be related but indicated as false predictions (“nonrelevant”). The parameter  $\beta \in [0, 1]$  indicates the relative influence of both measures.  $\beta \in [0, 1]$  which usually  $\beta = 1$  is considered.

Precision and F1, as decision support accuracy metrics, were utilized to assess the presented approach based on various values of Top-N. We considered  $N=5$  to  $N=100$  which indicates that we assessed the approach when suggesting the top 5 to 100 hotels by the proposed collaborative system. Table 9 shows the precision values for various Top-N. It is found that precision and F1 values obtained from our new approach are comparatively high in relation to the HOSVD + ANFIS [65], Pearson Nearest Neighbor, and SVD. This table also shows F1 measures for various Top-N. The results show that our approach has surpassed the Pearson Nearest Neighbor approach in all Top-N recommendations. For F1, the approach which utilized EM, HOSVD, and ANFIS works better than the Pearson Nearest Neighbor approach. It achieved  $F1 = 0.9112$  in the Top-100 recommendations. The advantage of our approach can be indicated by the usage of the fuzzy logic approach for the MC-CF part. These outcomes are adequate to endorse our hypothesis that the presented approach can present accurate and scalable outcomes in relation to the Pearson Nearest Neighbor approach.

We also assessed the approach based on various degrees of sparsity and measured the average MAE. Hence, we generate 10 sets of data with various sparsity degrees for the TripAdvisor sets of data (i.e., 99.5%, 98.5%, 97.5%, 96.5%, 95.5%, 94.5%, 93.5%, 92.5%, 91.5%, and 90.5%). We deployed the approach on the sets of data with these sparsity

TABLE 9: The results for precision and F1 of all approaches for different Top-N.

TOP-N	EM + HOSVD + ANFIS		EM + ANFIS		Pearson nearest neighbor		SVD		HOSVD + ANFIS	
	Precision	F-measure	Precision	F-measure	Precision	F-measure	Precision	F-measure	Precision	F-measure
Top5	0.7519	0.7647	0.6126	0.6363	0.4905	0.5403	0.5184	0.5675	0.6162	0.6447
Top10	0.7550	0.7729	0.6340	0.6464	0.5335	0.5490	0.5552	0.5762	0.6389	0.6472
Top15	0.7551	0.7748	0.6351	0.6899	0.5393	0.5586	0.5656	0.5843	0.6448	0.6962
Top20	0.7561	0.7909	0.6352	0.6988	0.5483	0.5795	0.5667	0.5967	0.6427	0.7013
Top25	0.7581	0.7939	0.6396	0.6994	0.5585	0.5857	0.5772	0.6032	0.6416	0.7067
Top30	0.7584	0.7975	0.6681	0.7183	0.5662	0.5944	0.5795	0.6223	0.6748	0.7248
Top35	0.7667	0.7981	0.6724	0.7200	0.5804	0.6037	0.5944	0.6507	0.6785	0.7201
Top40	0.7700	0.8005	0.6782	0.7239	0.6095	0.6353	0.6265	0.6567	0.6854	0.7294
Top45	0.7894	0.8042	0.6829	0.7240	0.6186	0.6403	0.6327	0.6597	0.6882	0.7246
Top50	0.7969	0.8091	0.6831	0.7249	0.6234	0.6491	0.6357	0.6631	0.6919	0.7291
Top55	0.7984	0.8293	0.6925	0.7304	0.6235	0.6496	0.6451	0.6758	0.7022	0.7339
Top60	0.8017	0.8331	0.6997	0.7321	0.6299	0.6546	0.6461	0.6768	0.7027	0.7337
Top65	0.8078	0.8343	0.7023	0.7356	0.6310	0.6663	0.6489	0.6768	0.7047	0.7375
Top70	0.8240	0.8395	0.7051	0.7477	0.6401	0.6672	0.6492	0.6796	0.7126	0.7497
Top75	0.8244	0.8563	0.7248	0.7541	0.6407	0.6701	0.6495	0.6814	0.7339	0.7549
Top80	0.8261	0.8603	0.7273	0.7638	0.6408	0.6710	0.6531	0.6827	0.7276	0.7711
Top85	0.8586	0.8870	0.7397	0.7655	0.6466	0.6737	0.6546	0.6843	0.7430	0.7709
Top90	0.8640	0.8974	0.7444	0.7657	0.6483	0.6848	0.6567	0.6881	0.7465	0.7726
Top95	0.8739	0.9009	0.7608	0.7723	0.6493	0.6855	0.6586	0.6906	0.7707	0.7807
Top100	0.8950	0.9112	0.7703	0.7835	0.6534	0.6878	0.6599	0.6996	0.7764	0.7889

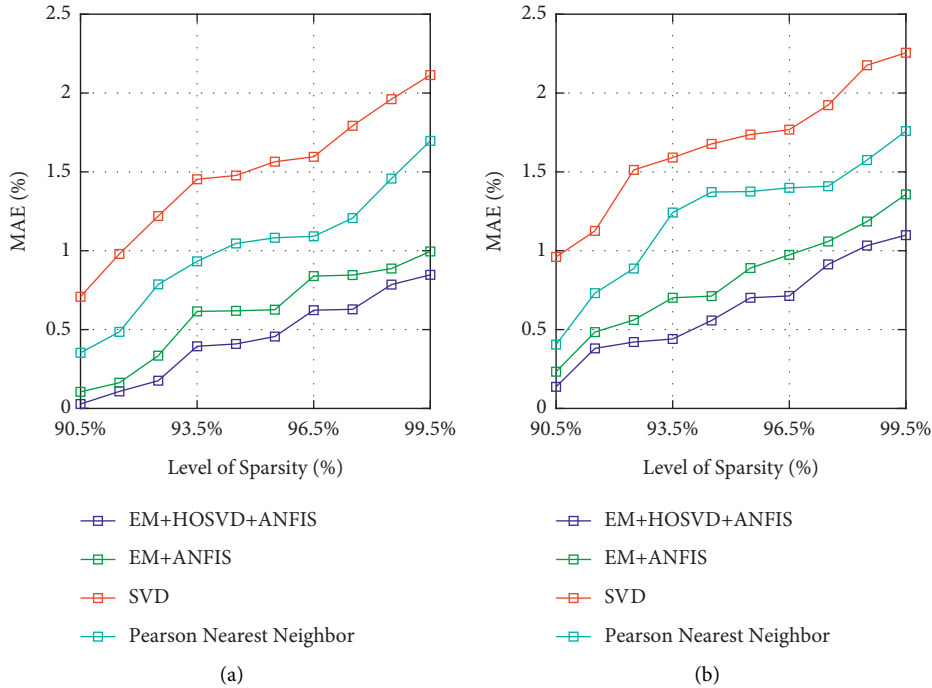


FIGURE 7: MAE and different data sparsity levels. (a) Neighbor set size = 20; (b) neighbor set size = 10.

degrees and assessed the outcomes compared to other RS techniques. Referring to Figures 7(a) and 7(b), the MAE values of the presented approach for all sparsity degrees of the dataset are less than other approaches. Besides, the growing ratio of the MAE for the Pearson Nearest Neighbor approach is considered very high in comparison with other approaches. The outcomes also indicated that using clustering and reducing the dimensionality along with the neurofuzzy method presented better prediction accuracy in comparison with other approaches for the sparser dataset. Thus, the approach which used the fuzzy logic approach with the aid of EM, HOSVD, and ANFIS is more efficient in addressing the sparsity problem and presenting more accurate results.

## 8. Discussion and Research Implications

The impact of the current health outbreak caused a steep decline in the travel and tourism market [77, 78], leading to a decline in the rating ratio of travelers for tourism products. Both industry and research fields need to gather and analyze the available data to investigate, design, and deploy appropriate recovery policies in the tourism sector [79]. An appropriate recommender system can help tourists to overcome their uncertainty, particularly during the current crisis, regarding their travel plans and destination choice. However, CF recommendation agents suffer from serious data sparsity issue which broadly impacts the performance of RSs [43], particularly in the domains that have a limited average of rated items [80]. In fact, a major issue in CF recommendation agents is the data sparsity problem [81, 82], in which it is difficult for the system to present efficient suggestions to users due to the lack of information [6]. This

issue can arise when the consumer interacts with a limited portion of products in a specific application field. The influence of data sparsity on the quality of RS can be represented by two major folds [38]. First, when it is hard to locate the neighbors of a specific user in the domain of sparse data, particularly when there is a lack of historical data, a high degree of sparsity will restrict the system's ability to build reliable neighborhoods [83]. Sometimes, even though few neighbors are there, this issue can impact the accuracy of recommendations. Second, inadequate ratings presented by neighbors will impact the coverage of the presented suggestions and the ability of the deployed algorithm to present a novel and accurate list of recommendations. Unrated items and inactive travelers (in the context of TRSs) cannot be involved in the CF procedure.

ML techniques have been used in several studies in the context of RSs studies [67, 84]. ML techniques simulate people's learning process and enable the identification and gaining of new information and, accordingly, enhance the effectiveness of particular functions based on the resulting information [85]. ML has been used to address the sparsity problem in several studies and by adopting various techniques such as sparse Bayesian extreme learning machine [86], spectral coclustering [87], and KNN [83].

A considerable piece of research has therefore been performed both in the business and in academic community to advance efficient prediction models for multicriteria RSs [32, 70, 88, 89]. In the context of the current crisis of COVID-19, to the best of our knowledge, no previous study has been conducted to investigate the sparsity issue in travel recommendation agents. As the COVID-19 pandemic continues, the issues of how to manage the continuous influences on the tourism and hospitality sectors have



received much attention from researchers and business managers [90]. The tourism and hospitality sectors have been broadly influenced by travelers-generated media which are represented by online reviews and ratings [91]. Travelers often seek help on where to travel and what activities to do. Travelers usually look for suggestions presented by other travelers. Still, human capabilities are limited which may lead to difficulty in finding suitable suggestions that match their needs and wishes. Hence, during this critical situation, RSs can be used effectively as intelligent tools and supporting systems in aiding users to reach the right choice [92]. The deployed RSs should be flexible to address the continuous obstacles of the current crisis and meet the travelers' needs [91]. In this work, a new approach was presented to enhance the prediction outcomes of RSs, particularly to address the data sparsity problem during the current global crisis.

The research has several theoretical, practical, and methodological contributions. From a theoretical standpoint, this research presents one of the first attempts to overcome the sparsity problem during the emerging and novel crisis of COVID-19, which supports the current research in RSs literature. To meet the aim of this research, we presented a new recommendation method and evaluated it on the TripAdvisor dataset. Considering the practical view of point, the research can present research insights to be deployed in the design of travel RSs to overcome the sparsity problem, particularly due to the lack of data and uncertainty in the current pandemic. Enhancing the accuracy of RSs is one of the basic aims of deploying new recommendation techniques. Inadequate data is a critical obstacle when working with decision-making systems. Hence, from the methodological view of point, it is important to present new methods to overcome the sparsity problem in the RSs. As indicated by the research outcomes, the deployed techniques, dimensionality reduction, and clustering, as well as neurofuzzy, presented a robust performance in addressing the sparsity problem. The deployed method can be used effectively to face the issue of inefficient recommendations when there are a lower number of ratings.

## 9. Conclusion

In this study, we attempted to solve the sparsity problem and to improve the performance of RSs in the tourism context. The aim was to provide a method to solve the sparsity issue in tourism datasets that can effectively help the travel recommendation agents in the current outbreak. The approach presented in this research utilizes dimensionality reduction and clustering techniques with the aid of supervised machine learning. The method was assessed through a dataset from TripAdvisor using precision, F1, and MAE measures. We compared our results with the previous algorithms. The outcomes presented by precision, F1, and MAE indicated that the use of clustering, dimensionality reduction, and neurofuzzy approaches was efficient in enhancing the performance of the MC-CF. The outcomes indicated that the hybrid recommendation approach can be utilized to address the sparsity problems of RSs in the tourism context. In this research, EM clustering was utilized for clustering and

nonincremental ANFIS and HOSVD were used for predicting and reducing the dimensionality of the data in tensors. Incremental ANFIS and HOSVD can aid the RS to present more scalable recommendations compared with the nonincremental HOSVD and SVD. Besides, in future research, we aim to investigate the integration of other clustering approaches with ensemble learning approaches to the proposed method. In addition, in this research, the proposed approach was assessed by focusing on the hotel domain. Future research can consider other domains such as e-commerce RSs. We also aim to further enhance the suggested approach and assess it by utilizing additional measures such as diversity and novelty.

## Data Availability

Data are available on request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was made possible by a generous fund from the deanship of scientific research at Taif University, Taif, Saudi Arabia, under Taif University Researchers Supporting Project, Project no. TURSP-2020/344. This research was supported by Princess Nourah Bint Abdulrahman University Researchers Supporting Project no. PNURSP2022R4, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.

## References

- [1] W. Husain and L. Y. Dih, "A framework of a personalized location-based traveler recommendation system in mobile application," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 7, no. 3, pp. 11–18, 2012.
- [2] A. Levi, O. Mokryn, C. Diot, and N. Taft, "Finding a needle in a haystack of reviews: cold start context-based hotel recommender system," in *Proceedings of the sixth ACM conference on Recommender systems*, pp. 115–122, Association for Computing Machinery, New York, NY, USA, September 2012.
- [3] L. Esmaili, S. Mardani, S. A. H. Golpayegani, and Z. Z. Madar, "A novel tourism recommender system in the context of social commerce," *Expert Systems with Applications*, vol. 149, Article ID 113301, 2020.
- [4] M. Nilashi, E. Yadegaridehkordi, O. Ibrahim, S. Samad, A. Ahani, and L. Sanzogni, "Analysis of travellers' online reviews in social networking sites using fuzzy logic approach," *International Journal of Fuzzy Systems*, vol. 21, no. 5, pp. 1367–1378, 2019.
- [5] M. Nilashi, A. Ahani, M. D. Esfahani et al., "Preference learning for eco-friendly hotels recommendation: a multi-criteria collaborative filtering approach," *Journal of Cleaner Production*, vol. 215, pp. 767–783, 2019.
- [6] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using

- linked open data,” *Expert Systems with Applications*, vol. 149, Article ID 113248, 2020.
- [7] G. Guo, J. Zhang, and D. Thalmann, “Merging trust in collaborative filtering to alleviate data sparsity and cold start,” *Knowledge-Based Systems*, vol. 57, pp. 57–68, 2014.
  - [8] M. Polato and F. Aioli, “Exploiting sparsity to build efficient kernel based collaborative filtering for top-N item recommendation,” *Neurocomputing*, vol. 268, pp. 17–26, 2017.
  - [9] H. Koochi and K. Kiani, “Two new collaborative filtering approaches to solve the sparsity problem,” *Cluster Computing*, vol. 24, no. 2, pp. 753–765, 2021.
  - [10] N. A. Osman, S. A. Mohd Noah, M. Darwich, and M. Mohd, “Integrating contextual sentiment analysis in collaborative recommender systems,” *PLoS One*, vol. 16, no. 3, Article ID e0248695, 2021.
  - [11] S. S. Choudhury, S. N. Mohanty, and A. K. Jagadev, “Multimodal trust based recommender system with machine learning approaches for movie recommendation,” *International Journal of Information Technology*, vol. 13, no. 2, pp. 475–482, 2021.
  - [12] A. Hoque, F. A. Shikha, M. W. Hasanat, I. Arif, and A. B. A. Hamid, “The effect of Coronavirus (COVID-19) in the tourism industry in China,” *Asian Journal of Multidisciplinary Studies*, vol. 3, no. 1, pp. 52–58, 2020.
  - [13] P. Jones and D. Comfort, “The COVID-19 Crisis and Sustainability in the Hospitality Industry,” *International journal of contemporary hospitality management*, vol. 32, 2020.
  - [14] J. Serrano-Guerrero, F. P. Romero, and J. A. Olivas, “Fuzzy Logic Applied to Opinion Mining: A Review,” *Knowledge-Based Systems*, vol. 222, Article ID 107018, 2021.
  - [15] L. A. Zadeh, *Fuzzy Logic, Neural Networks, and Soft Computing, Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh*, pp. 775–782, World Scientific, New York, NY, USA, 1996.
  - [16] I. Huitzil, F. Alegre, and F. Bobillo, “GimmeHop: a recommender system for mobile devices using ontology reasoners and fuzzy logic,” *Fuzzy Sets and Systems*, vol. 401, pp. 55–77, 2020.
  - [17] C. Porcel, C. Martinez-Cruz, J. Bernabé-Moreno, and A. Tejada-Lorente, “Integrating ontologies and fuzzy logic to represent user-trustworthiness in recommender systems,” *Procedia Computer Science*, vol. 55, pp. 603–612, 2015.
  - [18] R. Colomo-Palacios, I. González-Carrasco, J. L. López-Cuadrado, A. García-Crespo, and ReSySTER, “A hybrid recommender system for Scrum team roles based on fuzzy and rough sets,” *International Journal of Applied Mathematics and Computer Science*, vol. 22, pp. 801–816, 2012.
  - [19] M. Ferreira-Satler, F. P. Romero, V. H. Menendez-Dominguez, A. Zapata, and M. E. Prieto, “Fuzzy ontologies-based user profiles applied to enhance e-learning activities,” *Soft Computing*, vol. 16, no. 7, pp. 1129–1141, 2012.
  - [20] E. Majd and V. Balakrishnan, “A trust model for recommender agent systems,” *Soft Computing*, vol. 21, no. 2, pp. 417–433, 2017.
  - [21] A. Jain and C. Gupta, *Fuzzy Logic in Recommender Systems, Fuzzy Logic Augmentation of Neural and Optimization Algorithms: Theoretical Aspects and Real Applications*, pp. 255–273, Springer, New York, NY, USA, 2018.
  - [22] T. C. K. Huang, “Recommendations of closed consensus temporal patterns by group decision making,” *Knowledge-Based Systems*, vol. 54, pp. 318–328, 2013.
  - [23] C. Birtolo and D. Ronca, “Advances in clustering collaborative filtering by means of fuzzy C-means and trust,” *Expert Systems with Applications*, vol. 40, no. 17, pp. 6997–7009, 2013.
  - [24] V. Kant and K. K. Bharadwaj, “Integrating collaborative and reclusive methods for effective recommendations: a fuzzy Bayesian approach,” *International Journal of Intelligent Systems*, vol. 28, no. 11, pp. 1099–1123, 2013.
  - [25] J. Serrano-Guerrero, F. P. Romero, J. A. Olivas, and Hiperion, “A fuzzy approach for recommending educational activities based on the acquisition of competences,” *Information Sciences*, vol. 248, pp. 114–129, 2013.
  - [26] G. Castellano, M. G. Cimino, A. M. Fanelli, B. Lazzerini, F. Marcelloni, and M. A. Torsello, “A collaborative situation-aware scheme based on an emergent paradigm for mobile resource recommenders,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 4, pp. 421–437, 2013.
  - [27] J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, “A hybrid recommendation approach for a tourism system,” *Expert Systems with Applications*, vol. 40, no. 9, pp. 3532–3550, 2013.
  - [28] P. Paranjape-Voditel and U. Deshpande, “A stock market portfolio recommender system based on association rule mining,” *Applied Soft Computing*, vol. 13, no. 2, pp. 1055–1063, 2013.
  - [29] I. Wu and W. H. Hwang, “A genre-based fuzzy inference approach for effective filtering of movies,” *Intelligent Data Analysis*, vol. 17, no. 6, pp. 1093–1113, 2013.
  - [30] S. R. d M. Queiroz, F. d A. de Carvalho, G. L. Ramalho, and V. Corruble, *Making Recommendations for Groups Using Collaborative Filtering and Fuzzy Majority*, *Brazilian Symposium on Artificial Intelligence*, pp. 248–258, Springer, New York, NY, USA, 2002.
  - [31] L. Martínez, M. J. Barranco, L. G. Pérez, and M. Espinilla, “A knowledge based recommender system with multigranular linguistic information,” *International Journal of Computational Intelligence Systems*, vol. 1, no. 3, pp. 225–236, 2008.
  - [32] M. Nilashi, O. bin Ibrahim, and N. Ithnin, “Hybrid recommendation approaches for multi-criteria collaborative filtering,” *Expert Systems with Applications*, vol. 41, no. 8, pp. 3879–3900, 2014.
  - [33] J. Bobadilla, A. Hernando, F. Ortega, and J. Bernal, “A framework for collaborative filtering recommender systems,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 14609–14623, 2011.
  - [34] A. Taneja and A. Arora, “Cross domain recommendation using multidimensional tensor factorization,” *Expert Systems with Applications*, vol. 92, pp. 304–316, 2018.
  - [35] D. Anand and K. K. Bharadwaj, “Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 5101–5109, 2011.
  - [36] Z. D. Champiri, S. R. Shahamiri, and S. S. B. Salim, “A systematic review of scholar context-aware recommender systems,” *Expert Systems with Applications*, vol. 42, no. 3, pp. 1743–1758, 2015.
  - [37] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
  - [38] F. Zhang, S. Qi, Q. Liu, M. Mao, and A. Zeng, “Alleviating the data sparsity problem of recommender systems by clustering nodes in bipartite networks,” *Expert Systems with Applications*, vol. 149, Article ID 113346, 2020.
  - [39] G. Guo, J. Zhang, and N. Yorke-Smith, “Trustsvd: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings,” in *Proceedings of the AAAI*

- Conference on Artificial Intelligence*, Association for Computing Machinery, New York, NY, USA, January 2015.
- [40] J. Tang, S. Wang, X. Hu et al., "Recommendation with social dimensions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, California, CA, USA, February 2016.
- [41] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using bayesian personalized ranking for social network data," in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, pp. 173–182, Association for Computing Machinery, New York, NY, USA, February 2012.
- [42] T. Zhao, J. McAuley, and I. King, "Leveraging social connections to improve personalized ranking for collaborative filtering," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pp. 261–270, Association for Computing Machinery, New York, NY, USA, November 2014.
- [43] C. Feng, J. Liang, P. Song, and Z. Wang, "A fusion collaborative filtering method for sparse data in recommender systems," *Information Sciences*, vol. 521, pp. 365–379, 2020.
- [44] G. Guo, H. Qiu, Z. Tan, Y. Liu, J. Ma, and X. Wang, "Resolving data sparsity by multi-type auxiliary implicit feedback for recommender systems," *Knowledge-Based Systems*, vol. 138, pp. 202–207, 2017.
- [45] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 165–172, Association for Computing Machinery, New York, NY, USA, October 2013.
- [46] X. Zhu, Z. Y. Ming, Y. Hao, and X. Zhu, "Tackling data sparseness in recommendation using social media based topic hierarchy modeling," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, AAAI Press, California, CA, USA, July 2015.
- [47] R. He and J. McAuley, "VBPR: visual bayesian personalized ranking from implicit feedback," in *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI Press, California, CA, USA, February 2016.
- [48] F. Vasile, E. Smirnova, and A. Conneau, "Meta-prod2vec: product embeddings using side-information for recommendation," in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 225–232, Association for Computing Machinery, New York, NY, USA, September 2016.
- [49] W. Pan, H. Zhong, C. Xu, and Z. Ming, "Adaptive Bayesian personalized ranking for heterogeneous implicit feedbacks," *Knowledge-Based Systems*, vol. 73, pp. 173–180, 2015.
- [50] H. Qiu, G. Guo, J. Zhang, Z. Sun, H. T. Nguyen, and Y. Liu, "Tbpr: trinity preference based bayesian personalized ranking for multivariate implicit feedback," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, pp. 305–306, Association for Computing Machinery, New York, NY, USA, July 2016.
- [51] B. Loni, R. Pagano, M. Larson, and A. Hanjalic, "Bayesian personalized ranking with multi-channel user feedback," in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 361–364, Association for Computing Machinery, New York, NY, USA, September 2016.
- [52] R. A. Hamid, A. Albahri, J. K. Alwan et al., "How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management," *Computer Science Review*, vol. 39, Article ID 100337, 2021.
- [53] N. Leiper, *Tourism Systems: An Interdisciplinary Perspective*, Department of Management Systems, Business Studies Faculty, New Zealand, Oceania, 1990.
- [54] S. L. Smith, *Tourism Analysis: A Handbook*, Routledge, England, UK, 2014.
- [55] T. Y. Chou, C. L. Hsu, and M. C. Chen, "A fuzzy multi-criteria decision model for international tourist hotels location selection," *International Journal of Hospitality Management*, vol. 27, no. 2, pp. 293–301, 2008.
- [56] J. M. Benitez, J. C. Martin, and C. Román, "Using fuzzy number for measuring quality of service in the hotel industry," *Tourism Management*, vol. 28, no. 2, pp. 544–555, 2007.
- [57] A. H. Busalim and A. R. C. Hussin, "Understanding social commerce: a systematic literature review and directions for further research," *International Journal of Information Management*, vol. 36, no. 6, pp. 1075–1088, 2016.
- [58] U. Gretzel and K. H. Yoo, "Use and Impact of Online Travel Reviews, Information and Communication Technologies in Tourism 2008," in *Proceedings of the International Conference in Innsbruck*, pp. 35–46, DBLP, Germany, January 2008.
- [59] H. Han, H. Xu, and H. Chen, "Social commerce: a systematic review and data synthesis," *Electronic Commerce Research and Applications*, vol. 30, pp. 38–50, 2018.
- [60] T. P. Liang, Y. T. Ho, Y. W. Li, and E. Turban, "What drives social commerce: the role of social support and relationship quality," *International Journal of Electronic Commerce*, vol. 16, no. 2, pp. 69–90, 2011.
- [61] F. Xu, D. Buhalis, and J. Weber, "Serious games and the gamification of tourism," *Tourism Management*, vol. 60, pp. 244–256, 2017.
- [62] D. Buhalis and A. Amaranggana, *Smart Tourism Destinations Enhancing Tourism Experience through Personalisation of Services, Information and Communication Technologies in Tourism*, pp. 377–389, Springer, New York, NY, USA, 2015.
- [63] N. Hajli, "Social commerce constructs and consumer's intention to buy," *International Journal of Information Management*, vol. 35, no. 2, pp. 183–191, 2015.
- [64] B. Lu, W. Fan, and M. Zhou, "Social presence, trust, and social commerce purchase intention: an empirical research," *Computers in Human Behavior*, vol. 56, pp. 225–237, 2016.
- [65] M. Nilashi, O. bin Ibrahim, N. Ithnin, and N. H. Sarmin, "A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA-ANFIS," *Electronic Commerce Research and Applications*, vol. 14, no. 6, pp. 542–562, 2015.
- [66] S. Loh, F. Lorenzi, R. Saldaña, and D. Lichnow, "A tourism recommender system based on collaboration and text analysis," *Information Technology & Tourism*, vol. 6, no. 3, pp. 157–165, 2003.
- [67] M. Nilashi, K. Bagherifard, M. Rahmani, and V. Rafe, "A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques," *Computers & Industrial Engineering*, vol. 109, pp. 357–368, 2017.
- [68] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
- [69] D. Jannach, Z. Karakaya, and F. Gedikli, "Accuracy improvements for multi-criteria recommender systems," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, pp. 674–689, Association for Computing Machinery, New York, NY, USA, June 2012.
- [70] M. Nilashi, D. Jannach, O. bin Ibrahim, and N. Ithnin, "Clustering-and regression-based multi-criteria collaborative filtering with incremental updates," *Information Sciences*, vol. 293, pp. 235–250, 2015.

- [71] G. Adomavicius and Y. Kwon, "New recommendation techniques for multicriteria rating systems," *IEEE Intelligent Systems*, vol. 22, no. 3, pp. 48–55, 2007.
- [72] M. Nilashi, O. Ibrahim, and K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques," *Expert Systems with Applications*, vol. 92, pp. 507–520, 2018.
- [73] S. Kant, T. Mahara, V. K. Jain, and D. K. Jain, "Fuzzy logic based similarity measure for multimedia contents recommendation," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4107–4130, 2019.
- [74] M. Nilashi, B. Minaei-Bidgoli, A. Alghamdi et al., "Knowledge discovery for course choice decision in Massive Open Online Courses using machine learning approaches," *Expert Systems with Applications*, vol. 199, Article ID 117092, 2022.
- [75] J. S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.
- [76] M. Ghavipour and M. R. Meybodi, "An adaptive fuzzy recommender system based on learning automata," *Electronic Commerce Research and Applications*, vol. 20, pp. 105–115, 2016.
- [77] J. Abbas, "Crisis Management, Transnational Healthcare Challenges and Opportunities: The Intersection of COVID-19 Pandemic and Global Mental Health," *Research in Globalization*, vol. 3, Article ID 100037, 2021.
- [78] E. Avery, "Contextual and audience moderators of channel selection and message reception of public health information in routine and crisis situations," *Journal of Public Relations Research*, vol. 22, no. 4, pp. 378–403, 2010.
- [79] M. Sigala, "Tourism and COVID-19: impacts and implications for advancing and resetting industry and research," *Journal of Business Research*, vol. 117, 2020.
- [80] X. Xu, K. Dutta, and C. Ge, "Do adjective features from user reviews address sparsity and transparency in recommender systems?" *Electronic Commerce Research and Applications*, vol. 29, pp. 113–123, 2018.
- [81] M. K. Najafabadi, A. Mohamed, and C. W. Onn, "An impact of time and item influencer in collaborative filtering recommendations using graph-based model," *Information Processing & Management*, vol. 56, no. 3, pp. 526–540, 2019.
- [82] Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 116–142, 2004.
- [83] M. Grčar, D. Mladenič, B. Fortuna, and M. Grobelnik, "Data sparsity issues in the collaborative filtering framework," in *Proceedings of the International Workshop on Knowledge Discovery on the Web*, pp. 58–76, Springer, New York, NY, USA, October 2005.
- [84] A. Yassine, L. Mohamed, and M. Al Achhab, "Intelligent recommender system based on unsupervised machine learning and demographic attributes," *Simulation Modelling Practice and Theory*, vol. 107, Article ID 102198, 2021.
- [85] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: a systematic review," *Expert Systems with Applications*, vol. 97, pp. 205–227, 2018.
- [86] J. Luo, C. M. Wong, and C. M. Vong, "Multinomial Bayesian extreme learning machine for sparse and accurate classification model," *Neurocomputing*, vol. 423, pp. 24–33, 2021.
- [87] M. Nasiri, B. Minaei, and Z. Sharifi, "Adjusting data sparsity problem using linear algebra and machine learning algorithm," *Applied Soft Computing*, vol. 61, pp. 1153–1159, 2017.
- [88] Q. Shambour, M. Hourani, and S. Fraihat, "An item-based multi-criteria collaborative filtering algorithm for personalized recommender systems," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 8, pp. 274–279, 2016.
- [89] Q. Shambour, "A user-based multi-criteria recommendation approach for personalized recommendations," *International Journal of Computer Science and Information Security*, vol. 14, no. 12, p. 657, 2016.
- [90] J. Yin and Y. Ni, "COVID-19 event strength, psychological safety, and avoidance coping behaviors for employees in the tourism industry," *Journal of Hospitality and Tourism Management*, vol. 47, 2021.
- [91] M. Nilashi, S. Asadi, R. A. Abumalloh, S. Samad, F. Ghabban, and A. Ahani, "Recommendation agents and information sharing through social media for coronavirus outbreak," *Telematics and Informatics*, vol. 61, Article ID 101597, 2021.
- [92] R. Abumalloh, O. Ibrahim, and M. Nilashi, "Loyalty of Young Female Arabic Customers towards Recommendation Agents: A New Model for B2C E-Commerce," *Technology in Society*, vol. 61, Article ID 101253, 2020.