RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

A Hybrid Method for Customer Segmentation in Saudi Arabia Restaurants Using Clustering, Neural Networks and Optimization Learning Techniques

Abdullah Alghamdi[1]

Abstract

Dining out is one of the biggest expenditures for travelers worldwide and is essential for tourism-dependent destinations. Market segmentation gives industries the potential to classify similar customers and categorize their preferred target markets to ensure marketing expenses' operative management. It has been a practical approach for business improvement in tourism and hospitality. Big data are fundamentally changing the management of the hospitality sector and the relationship between the customer and business by simplifying the decision-making process based on large amounts of data. The data provided in social media have played an important role in customer segmentation. In fact, the data provided by the customers in social media have been a valuable source for decision-makers to precisely discover the customers' satisfaction dimensions on their services. Therefore, there is a need for the development of data-driven approaches for social data analysis for customers segmentation. This research aims to develop a new data-driven approach to reveal customers' satisfaction in restaurants. Specifically, k-means and Artificial Neural Network (ANN) with the aid of the Particle Swarm Optimization (PSO) technique are, respectively, used in data clustering and prediction tasks. In this research, the data of customers on the service quality of restaurants are collected from the TripAdvisor platform. The results of the data analysis are provided. We evaluate the prediction model through a set of evaluation metrics, Mean Squared Error (MSE) and coefficient of determination (R^2), compared with the other prediction approaches. The results showed that k-means-PSO-ANN (MSE $= 0.09847$; $R^2 = 0.98764$) has outperformed other methods. The current study demonstrates that the use of online review data for customer segmentation can be an effective way in the restaurant industry in relation to the traditional data analysis approaches.

Keywords Neural Network · Optimization learning techniques · Customer decision-making · Data-driven analysis · Customer segmentation · Restaurants

1 Introduction

Dining out is one of the biggest expenditures for travelers around the world and is essential for tourism-dependent destinations [1]. Therefore, restaurants must determine what elements affect elevate spending and positive behavioral intentions to improve sales and word of mouth [2]. In the tourism and hospitality industry, travelers mainly trust the online setting to assess various products and services. Particularly, electronic Word-Of-Mouth (eWOM) has been considered as one of the essential references of data for customers [3]. Perhaps eWOM is necessary at the pre-trip phase when travelers determine their destination and most of the products they want to buy in their trip [3].

By using user-generated content, including the ability to share photographs, videos and comments with other customers, the researchers also found that social media can help the hospitality industry attract potential visitors, increase their income and thus lead to more success [4]. Today, user-generated content is considered by researchers and professionals in many industries as the most common and efficient type of marketing strategies tool [5]. Online reviews, published on tourism websites, such as TripAdvisor, Yelp and Expedia, have obtained significant attention from various theoretical or practical aspects [6].

✉ Abdullah Alghamdi
   aaalghamdi@nu.edu.sa

1 Information Systems Department, College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia

Springer

In this way, data analytics methods are vastly used to collect, summarize and interpret user-generated content to provide meaningful patterns and insights about management issues [7, 8]. Online review research has several advantages, such as data availability, clarity and simplicity of the data collection process. The mentioned data are often text-based and usually include huge data resources, or big data, that exceed the processing capacity of the conventional data analytics approaches [9].

Big data are fundamentally changing the management of the hospitality sector and the relationship between the customer and business, by simplifying the decision-making process based on large amounts of data. The hospitality sector has converted to an information-intensive industry where large volumes of data are stored with practical applications that are not widely wide-spread. By advent of big data, it is feasible to manipulate these data to obtain the determined goals and to convert the data to knowledge [3, 10, 11].

Market segmentation gives industries the potential to classify similar customers and categorize their preferred target markets, to assure the operative management of marketing expenses. From the hospitality and tourism perspective, it is essential to determine the correct behaviors and select various customer groups to create marketing strategies and shape customer interactions with the industry. Hotel managers should determine customers satisfaction and demands to enhance their marketing strategy and decision-making process [12, 13]. In other words, segmentation utilizing user-generated content is necessary if managers want to get a more idea of the unique preferences of customers and their level of satisfaction [14]. Interaction among big data and social media can reorganize the way that hospitality and tourism sectors interact with their customers [15]. Based on the literature, social media data are one of the most effective methods to shape consumer perspectives toward a product or service. Therefore, in this study, a hybrid machine learning approach utilizing neural networks and an optimization learning method for customer segmentation in Saudi Arabia restaurants was developed. Although clustering and neural networks are widely used in many prediction problems, the combination of these approaches in online reviews analysis is presented for the first time in this paper. In fact, the use of these approaches with the aid of optimization machine learning can still be novel for customer segmentation in Saudi Arabia restaurants.

This paper concentrates on the significance of online customers' reviews in revealing customers' satisfaction in restaurants. The current work mainly demonstrates the use of online review data for customer segmentation in the restaurant industry and compares its benefit contrasted with traditional data analysis approaches. We used clustering and neural network to develop the hybrid method. The performance of the neural network technique was improved using the optimization machine learning approach. The proposed method in this research highlights the role of customer segmentation in hospitality competitiveness. We investigate the preferences of restaurant customers based on food quality and service quality, and these results lead to recognizing customer clusters. Furthermore, the finding of this study will help managers to classify their customers into different clusters to deliver personalized services, intensifying customer gaining and inspiring excellent resource management in the hospitality industry. The current research develops novel market segmentation techniques utilizing big social media and online review platforms such as TripAdvisor. This paper also determines how managers can efficiently segment their customers through online customers' reviews. The provided hybrid machine learning technique can intensify the quality of segmentation analysis in a restaurant based on customer preferences and generated content.

This work is organized as follows. We present the literature review in Sect. 2. In Sect. 3, the hybrid method is presented. In Sect. 4, we provide the experimental results. Finally, in Sect. 5 and Sect. 6, the discussion and conclusion are, respectively, provided. The list of acronyms used in this paper is displayed in Table 1.

2 Research Background

2.1 Market Segmentation in Restaurant

User-generated data are one of the main sources of data related to customers, their needs and preferences. These data can explain the characteristics of users and their level of satisfaction with the product and services provided [16]. Customer segmentation is mainly defined as classifying customers into different groups. The users in each group have similar characteristics while differing from other groups. This method is helpful for industries to identify different customer groups [17]. Various customer segmentation methods have been used in the past researches. Customers with the same needs, such as shopping patterns and purchasing behavior, can be grouped into the same clusters through market segmentation. Based on this method, the customers are grouped utilizing their preferences and buying behavior. More accurate and correct customer segmentation in an industry will help organizational success. Market segmentation is primarily aimed at accurately predicting user needs, which will lead to increased profitability by investing or producing a suitable amount of goods within a certain period for final users at an appropriate cost [10]. Accordingly, organizations can communicate more efficiently with their customers, and users [18]. Customer segmentation is a common practice across many industries such as marketing [19], education [20] and tourism [21].

Table 1 List of abbreviations used in this manuscript

Abbreviation	Description
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANN	Artificial Neural Network
CART	Classification and Regression Tree
CHAID	Chi-Squared Automatic Interaction Detection
COVID-19	Coronavirus Disease 2019
CS	Customer Satisfaction
CSS	Customer-Related Social Stressors
XNN	Explainable Neural Network
eWOM	Electronic Word of Mouth
HOSVD	Higher-Order Singular Value Decomposition
KNN	K-Nearest Neighbors
LDA	Latent Dirichlet Analysis
MCDM	Multi-Criteria Decision-Making
MLR	Means-Multiple Linear Regression
MSE	Mean Squared Error
NOR	Negative Online Review
PF	Price Fairness
PFQ	Perceived Food Quality
PLTS	Probabilistic Linguistic Term Set
PV	Perceived Value
RFM	Recency, Frequency and Monetary
SOM	Self-Organizing Map
SVR	Support Vector Regression
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
EM	Expectation Maximization

The segmentation method in the context of tourism and hospitality focuses on identifying the requirements and benefits associated with specific product groups, user purchases/behavior, demographics and nationalities [22]. However, much of the research in customer segmentation has concentrated on applying Google Analytics dataset, yet, there is a growing effort to applying customer segmentation by social media data from the main online environments.

Some of the past research (see Table 2) considered restaurant customers segmentation in various dimensions such as demographic characteristics, preferences, behavior and satisfaction level. As an example, [23] developed an approach for spa hotel classification and determining travel choice utilizing data mining methods. Based on findings, 9 clusters for spa hotel segmentation were identified.

A meal at a restaurant or the purchase of a hotel room can also be considered a hospitality product [24]. Recently, there has been a growing interest among consumers in the use of vegetarian products [25]. The vegetarian concept is widely applied in a population that does not eat meat, seafood or foods made up of these elements [26]. Recent evidence suggests that vegetarian foods are an important source of a wide variety of essential micronutrients, and there is strong evidence that eating vegetarian can be effective in preventing many chronic diseases, such as cardiovascular disorders, diabetes and some cancers [27]. According to this, in a study conducted by [17], customer classification and determining their preferences in vegetarian restaurants were performed and customer preferences prediction was conducted through Classification and Regression Tree (CART) algorithm in different groups. In each group, a decision tree model is created to determine the customers' preferences. Furthermore, Trappey et al. [28] use preference variables to determine customers' loyalty to Japanese-style chain restaurants by K-means clustering. Based on findings, the restaurant created customized price discounts for each customer according to their past preferences and behaviors. In addition, profiling profitable hotel customers by Recency, Frequency and Monetary (RFM) indicators evaluated in [13]. In this method, Recency, Frequency and Monetary indicators are used for determining the type of customers.

A further study used a hybrid technique for user-generated data investigation by multi-criteria decision-making and text mining methods to determine the significance of elements pertinent to customers decision-making to green hotels selection [29]. Similarly in [30] classification of customer satisfaction in Las Vegas hotels was conducted. Based on results, naive Bayesian method can classify hotel customers with high precision and recall. Moreover, Nilashi et al. [10] provided an analytical approach for customer segmentation in eco-friendly hotels. In this research by utilizing the Expectation–Maximization (EM) method, 4 clusters were identified and travelers' satisfaction levels were determined. Furthermore, an investigation of the travelers' choice behavior in green hotels was conducted in [31]. The developed techniques can examine travelers' reviews and classifying eco-friendly hotels to determine the next choice behavior and assist customers in their decision-making.

Segmentation techniques can be categorized according to two main parameters, a-priori and post-hoc methods. In the a-priori approach, the type and amount of clusters are determined priori the data gathering. In the post hoc method, the number of groups and their attributes is set out by the current study approach [32]. Despite the popularity of post hoc methods, to date, rare researches have been advocated in the hospitality sector for spa hotel segmentation and customer preferences. Furthermore, despite the vast advancement in the tourism industry, only a limited number of papers concentrated on customers, preferences and satisfaction in vegetarian restaurants. Today big social data and user-generated contents create a large repository of data that demonstrates users' satisfaction from restaurant services quality. Furthermore, user-generated contents have

Table 2 A literature review of customer segmentation in restaurant

References	Study objective	Method	Results
[23]	Developing a model for spa hotel customers segmentation	Machine learning methods such as SOM, HOSVD, CART	The machine learning approach is a suitable method for spa hotel segmentation
[84]	Investigating the relationship among the user-generated data and brand in the restaurant	Panel data analysis	Based on results that the financial effects of positive contents diminished with the existence of the brand
[17]	Applying machine learning techniques for customer segmentation in vegetarian restaurants	Text mining, cluster analysis, and predictive learning technique	Based on findings customers' online reviews can be an effective tool to determine their preferences
[16]	Using probabilistic linguistic term set (PLTS) to determine the cluster analysis for restaurant selection	K-means clustering	A novel restaurant recommendation method by cluster analysis is proposed
[28]	Using preference variables to determine consumers loyalty	K-means clustering	By using clustering, managers can offer customized price discounts for users
[13]	Profiling profitable hotel customers by Recency, Frequency, and Monetary (RFM) indicators	RFM analysis	RFM indicators can classify customers based on preferences
[29]	Proposing a hybrid method for online reviews analysis	Cluster analysis, text mining, and Neuro-Fuzzy technique	The impact of green hotels on customer satisfaction is investigated for different
[30]	Traveling groups		
[10]	Determining guest experience to increase hospitality services	Identifying pertinent attributes to determine customer satisfaction	The naive Bayesian method is a suitable method for review classification
[31]	Determining eco-friendly hotels by machine learning methods	MCDM method	The proposed technique is useful for customers segmentation

the potential to explain differences among customers with various preferences. Accordingly, online reviews create a beneficial means for a data-driven approach and post hoc market segmentation that provide insight into how managers can efficiently segment different customers.

2.2 Relationship Between Restaurant Experience and Online Review

User-generated data in social media are becoming the main sources of information for hospitality research and practices [9]. Nowadays, with an enormous amount of social media data and the vast amount of online reviews, it is difficult to manipulate customers' opinions [33]. These data sources can create real information about the restaurant industry. User-generated data are a type of e-WOM which is considered as the task of transferring information related to a service from one person to another by electronic tools [34].

Many studies have been done to determine the impact of online reviews in the hospitality and tourism industry. Many online review tools collect user-generated data and publish content to potential users. Each social media platform has exclusive characteristics; these tools permit consumers to post their opinion and ask them to appraise experiences based on a scale. There is a different platform for online reviews, such as Expedia, Yelp, AirBnB and TripAdvisor [8, 35]. Researches applying for online reviews often use a set of reviews and pertinent data, to obtain characteristics that permit the researcher to identify, explain or predict patterns that are useful from a theoretical and practical standpoint. This research complements traditional methods which rely primarily on surveys and interviews and demonstrates a promising research direction by using big social media data [18, 36].

TripAdvisor, as one of the favored online review platforms, provides travelers free online travel advice. This service now enables consumers from various market segments to search for restaurants utilizing their websites and mobile applications. Previous literature (see Table 3) concentrated on the online review opportunities in the hotel industry to obtain practical knowledge on customers' experiences, satisfaction level, and preferences [23, 37–41]. Based on these studies, three main factors for evaluating hospitality experience are: food quality, service quality and atmosphere [37, 42–46].

Researchers rely on nutritional characteristics to judge the quality of food, for instance, in a study conducted by [44] food quality, price, perceived value and customer satisfaction were essential for users revisit and positive WOM in organic food restaurants. In [46], it is evident that food

Table 3 Online review study in restaurant

References	Objective	Results
[45]	The effect of COVID-19 on hospitality services by LDA, EM, and ANFIS	COVID-19 has a profound impact on hospitality performance and overall customers satisfaction level
[85]	Sentiment-analysis method in Yelp reviews	Customers provided positive content than negative content in restaurant reviews
[86]	The relationship between user experience online reviews posting time	There is a reverse association between user experience and content posting time
[14]	MCDM for determining the satisfaction of travelers	Based on the results various users have different satisfaction levels and preferences
[23]	Machine learning techniques in eco-friendly hotels review	The findings demonstrate that the hybrid methods are useful in preference analysis
[51]	Combining social media data and historical sales data	The proposed method is useful for customer segmentation
[87]	Determining consumers' preferences for tourism products	Customer segmentation is a valuable tool that aids users' decision-making
[40]	Data mining model in TripAdvisor for hotel rating	Users' experience has a key role in hotel rating and creating positive word of mouth
[44]	Evaluating the role of PFQ, PF, PV, and Customer Satisfaction on customers revisit	Based on the results PFQ has a positive effect on PF and PV
[39]	Examining the influence of online review on restaurant performance	Based on the results, the number of online reviews has a positive effect on restaurant performance
[46]	Examining the role of eWOM on the popularity of hospitality	Based on findings, quality of food, environment, and service have a positive impact on the popularity of restaurants
[43]	Determining contributing factors on the hotel rating	Food, service, environment price, and ambiance are the main features determining star ratings
[42]	Identifying main factors in positive eWOM	The results have shown that physical setting has a profound impact on positive word of mouth
[88]	Investigating hotel loyalty programs in flyertalk.com members	The main clusters that were obtained were program experience, value, and process
[37]	Using Yelp reviews for restaurants categorization	Ingredient, type of food, taste, price, and hygiene are important in restaurant segmentation
[33]	Proposing a new method to determine comparative relations from online reviews	The proposed method enables the identification of top competitors using dissimilarity indicators
[89]	Analyzing the effects of traditional face-to-face customer-related social stressors (CSSs), and negative online review (NOR) stressor	The proposed method demonstrates that, by considering the contribution of CSSs, receipt of NORs predicts anger, and anger moderates the association among NOR receipt and two indices of burnout
[90]	Estimating the effect of different elements for online consumer review	The findings explain the impact of online consumer review elements on restaurant popularity
[91]	Examining how various service experiences lead to different intentions to create eWOM	Findings reveal that expert users have more motivation to create eWOM after negative experiences
[41]	Online restaurant reviews analysis in Yelp.com	There are different factors for certified green and non-certified restaurants rating
[11]	Applying the bootstrap resampling method in the international hotel chain	A deep relationship was established on the most representative attributes of repeaters being traveled without children
[39]	Analyzing the online reviews from different social media	Results demonstrated that the number of online reviews has a positive effect on hotel performance
[35]	Examining information quality of three online review platforms	The findings show that there is vast inconsistency in the illustration of the hospitality industry on these tools
[38]	Using machine learning models to analyzing reviews hotels of Tehran on TripAdvisor	KNN algorithm is the best approach for customer segmentation

quality, restaurant environment and services, as well as the amount of consumer-generated reviews, have a positive impact on the popularity of restaurants. Another research [39] reached the same conclusions. Furthermore, [43] added a few other features to improve restaurant performance by online reviews,based on results food, service and environment are the main factors contributing to star ratings, followed by cost and atmosphere. In addition, some of the other factors such as ingredients, taste, well-being and hygiene had a paramount impact on user satisfaction [37]. In addition, customers not only assess the food quality but also service quality through their dining experience [1]. The influence of service quality on user satisfaction and behavioral intentions, such WOM, has been demonstrated by several studies. For example, Nilashi et al. [45] demonstrated that service quality has an important impact on hotel performance and customers' satisfaction. Based on the result of another study, the physical environment had the most impact on the sharing of positive content by consumers [42].

2.3 Machine Learning in Market Segmentation

Big social data analytics is considered a new research agenda that uses a variety of data repositories and analytical methods to draw conclusions and predict real situations [35]. In particular, with the advancement of machine learning techniques, textual contents in social media provide a vast shared cognitive context and, so, have been interpreted in most of the application areas [35]. Customer segmentation is considered as a classification of users into different groups. The individuals in each group possess similar characteristics while differing from other groups. User segmentation is a practical approach that helps industries to recognize different clusters of users [47], and various customer segmentation approaches have been applied the in literature. For instance, in the post hoc market segmentation which is according to analyzing customer features, several methods have been introduced in the literature, such as Chi-Squared Automatic Interaction Detection (CHAID), Self-Organizing Map (SOM) and Artificial Neural Network (ANN). Furthermore, in some of the reviewed literature, the K-means clustering method was used for customer segmentation [16, 28] Other scholars applied hybrid machine learning techniques for clustering, for instance, Ahani et al. [23] developed a method for spa hotel segmentation utilizing SOM, HOSVD, CART methods. Moreover, the clustering approach by the self-organizing map and predictive methods were applied for segmentation and predict customers preferences in vegetarian restaurants [17]. In addition, a combination approach for user-generated data analysis by multi-criteria decision-making and text analysis methods to determine the significance of elements influence on the travelers decision-making in spa hotels was developed in [29].

In the past decade, social media has become a workable and important environment in the tourism and hospitality industry. New social media create better communication among organizations and consumers. This interaction can occur by applying different methods such as liking, commenting on or sharing user-generated content [48]. The main direction of current studies concentrated on user-generated content is defined as "Big Data" that can be applied to get deeper insights, offer high value to customers and gain a competitive advantage in hospitality and tourism [11]. The data obtained from these resources create some of the methods for customers segmentation, by cultural traits, needs and demographic features. Most tourism settings use a variety of basic characteristics for customer segmentation, the most prevalent of these features are demographic, socioeconomic and psychological attributes. Furthermore, segmentation restricts competition to the groups, since companies concentrate merely on the target cluster of their preferences rather than the whole market. Correct customer segmentation also benefits the marketing strategies, enabling marketing managers to concentrate on the creating of novel messages and furthermore effectively distributing them through the most suitable methods. In other words, segmentation allows that positive word of mouth to be generated and disseminated to members of each specific group [49].

Concerning methodology, "priori" and "post hoc" are critical methods in market segmentation. The first method relates to the selection of features at the first step, subsequently the segmentation of the target group, while the second acts by the gathering of data; according to the correlated features, prior the customers are grouped into clusters with vast most similarities [3, 19]. For segmentation features, the first and most prevalent type of user clustering is by demographic traits, such as age, gender, language that generally select by managers for the clustering of users into different groups [50]. To deviate from these primary characteristics, the hospitality industry has started to use behavioral and psychological features for customer segmentation [18, 21]. This high amount of data necessitates the applying of advanced statistical methods. In the reviewed literature for customer segmentation, clustering seems to be the most prevalent approach for use. Today, data-driven methods are constantly growing, mainly by using tools such as web crawling, machine learning and other analytical methods for the aggregation and interpretation of big social data [3, 51]. In big social data analysis, a mixture of various machine learning methods seems to be a practical approach in obtaining appropriate segmentation outcome in both computational and marketing contexts.

3 Methodology

In this research, we aimed to develop a new data-driven approach to reveal customers' satisfaction in restaurants. The proposed method is shown in Fig. 1. As seen from this figure, we use clustering and neural network for customers segmentation through social data in social networking sites. We use clustering and neural network for analyzing customers' satisfaction. Specifically, k-means and neural network with the aid of particle swarm optimization technique are, respectively, used in data clustering and prediction tasks. After clustering the data, neural network is applied on the clusters for the prediction of customers' satisfaction. We evaluate the prediction models through a set of evaluation metrics, Mean Squared Error (MSE) [52] and coefficient of determination (R^2) [53]. Finally, comparisons are made to show the robustness of the proposed method in relation to the previous methods, Means-Multiple Linear Regression (MLR), Support Vector Regression (SVR) and neural network.

3.1 Clustering Using K-Means

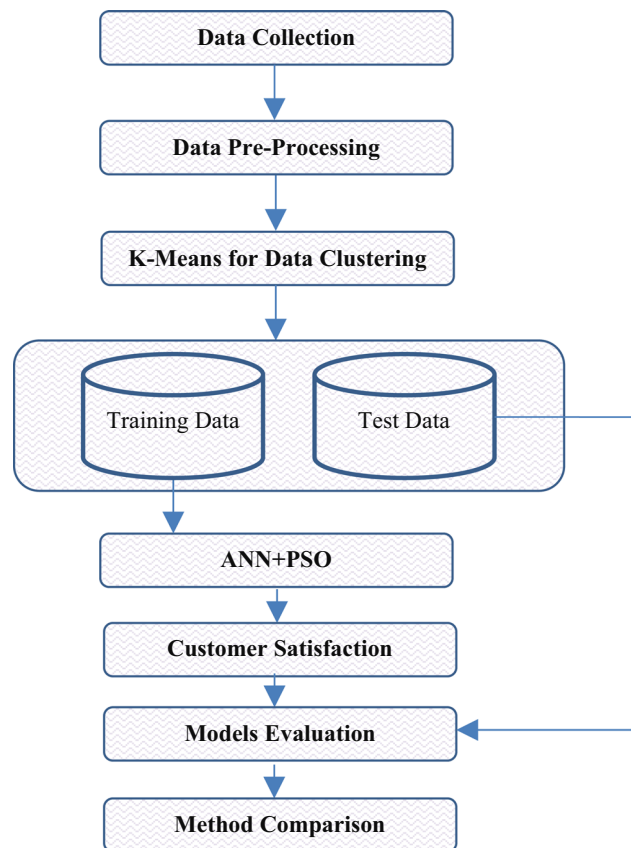K-means has been an effective clustering technique which has been widely used knowledge discovery [54–57]. This clustering algorithm divides data points into k clusters. These clusters, $S_i (i = 1, 2, \ldots, k)$, are presented by their own representative (cluster center) C_i. $S = \{\mathbf{X}\}$ represents the set of data points. For tow vectors $\mathbf{X}$ and $\mathbf{Y}$, the Euclidean distance of these two vectors is defined as $d(\mathbf{X}, \mathbf{Y})$. The main process of k-means clustering is shown in Algorithm 1.

Algorithm 1: The procedure of the K-means algorithm

Inputs: X: A dataset with N points; k: Arbitrary number of clusters
Output: C = $\{c_1, \ldots, c_k\}$: A set of k centroids for cluster k clusters
Procedure
C ← RandomSample(X, k)
Repeat
1. $S_i \leftarrow \{\text{x}: ||\text{x} - c_i||^2 \leq ||\text{x} - c_j||^2 \forall j, 1 \leq i, j \leq k\}$ such that $S_i \cap S_j = \emptyset$
2. $c_i \leftarrow \frac{1}{|s_i|} \Sigma_{x_j \in S_i} x_j$
Until assignments S_j do not change
Return C
End Procedure

3.2 ANN

As a computational approach, Artificial Neural Network (ANN) uses a directed network of linked neurons to map inputs to outputs [58, 59]. Each neuron in the network is a fundamental computation unit that computes $y = max\left(0, \sum_i w_i x_i + b\right)$, where $\{x_i\}$ denotes the neuron inputs, $\{w_i\}$ denotes the neuron weights, b denotes the neuron bias, and y indicates the neuron output. Each neuron is coupled in a layered architecture, with the following formula used to map inputs to outputs:

$$h_i = max(0, W_i \cdot h_{i-1} + b_i) \text{ for } 1 \leq i \leq L \text{ and } h_0 = x \tag{1}$$

$$y = max(0, V h_L) \tag{2}$$

where V, vector $b_1, \ldots, b_L$ and matrices $W_1, \ldots, W_L$ are parameters of the model and L indicates the number of layers. The parameters are learnt from the dataset.

The data collected were used to train the artificial neural network. The artificial neural network inputs were the features of the dataset, and its output was the level of customer satisfaction. The entire dataset was divided into two parts: a training set (80% of the data) and a testing set (20% of the data). The validation method used is tenfold cross-validation [60]. Prior to application, PSO tuned the architecture of the ANN based on the MSE according to the following equation:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i^* - y_i)^2 \tag{3}$$

Fig. 1 Proposed hybrid method

where N indicates the number of instances; y_i^* and y_i denote, respectively, the predicted and experimental customers' satisfaction levels of the ith instance in the dataset.

A nonlinear function of the inputs is obtained by using an ANN, and the ANN outputs are controlled by weights obtained in the learning process in the network. The learning process and learning approach are mainly supervised and Back Propagation (BP). It is necessary to have a bounded differentiable activation function for the BP training process to work properly. The sigmoid function, which is the most well known of all functions, has been employed. It is restricted to the range between the minimum (0) and maximum (1). To pass the signal to the next layer of neurons, the neuron summed output is scaled by sigmoid function prior to being passed on. A network ability to convert the weights in the network in response to errors is essential for error propagation network learning. In Fig. 2, we present the schematic diagram of the learning algorithm use in the paper.

3.3 PSO

Particle Swarm Optimization (PSO) is a highly effective optimization approach for locating the global optimum in a multidimensional search space [61–63]. The learning process of PSO began with swarm of particles which is randomly generated, each of which represents a distinct architecture of an ANN. To determine the fitness of particles' positions on the training set, the MSE metric is mainly used. For a more accurate network, architecture of an artificial neural network with a lower mean squared error is represented by a higher fitness particle. Particle position updates are used to generate next swarm, which took into account each particle' best position in history and swarm's best position in history. Swarms of particles were developed to the optimal position incrementally until the network reaches a maximum iteration. The following formulas are used in this paper to determine the position of particles:

$$V_i^{r+1} = wV_i^r + c_1r_1\left(p_{best.i}^r - X_i^r\right) + c_2r_2\left(g_{besf.i}^r - X_i^f\right) \quad (4)$$

$$X_i^{f+1} = X_i^r + V_i^{f+1} \quad (5)$$

where w denotes the inertia parameter. The two parameters c_1 and c_2 indicate, respectively, the cognitive and social influences; V_i^{t+1} and V_i^r are, respectively, the velocity of particle i at r and $r+1$ iterations ; X_i^{r+1} and χ_i^r indicate positions of particle$i$; r_1 and r_2 indicate random values in the ranges of 0 to 1; $g_{besf.i}^r$ and p_{besri}^r are, respectively, the swarm best position and the best position of a particle. The procedure for PSO is presented in Fig. 3.
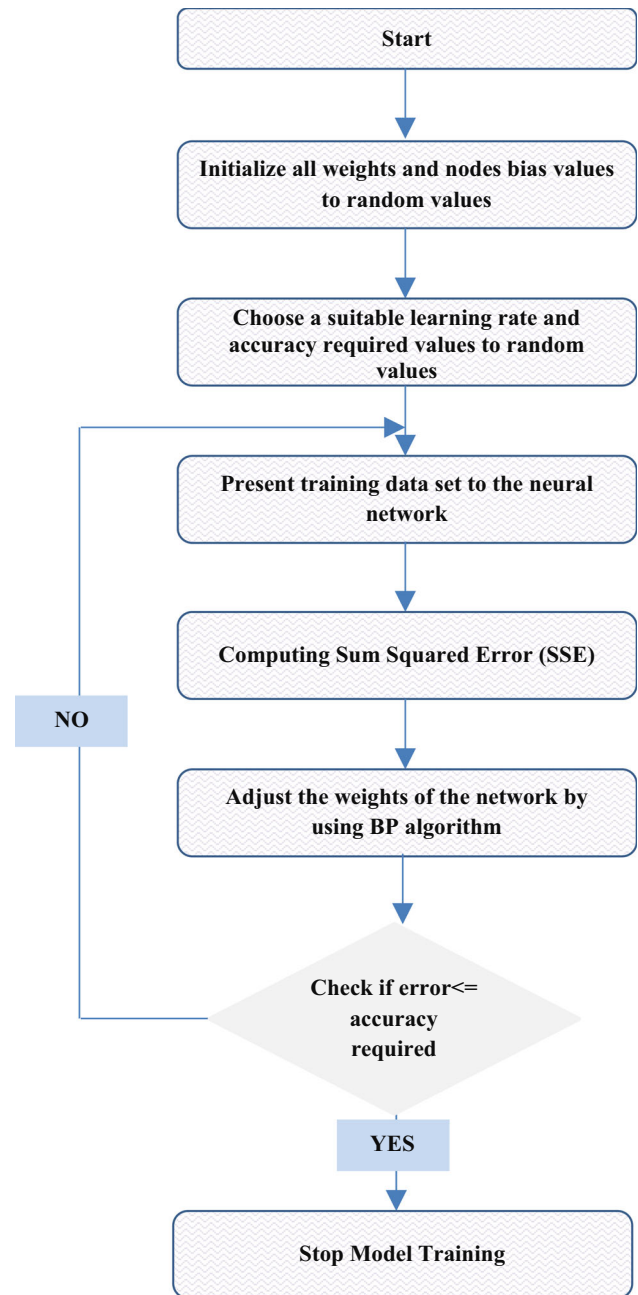

Fig. 2 Constructing prediction models in ANN using PBP

4 Data Collection and Method Evaluation

To evaluate the method, we collected a set of data from TripAdvisor. The Saudi Arabia restaurants registered in TripAdvisor are selected for data collection (see Fig. 4). The numerical ratings were collected and analyzed by the proposed method.

In TripAdvisor, the customers provide ratings in a range of 1 to 5 on the restaurants. There are four main criteria (Food, Service, Value and Atmosphere) to assess the quality of a

Fig. 3 PSO procedure
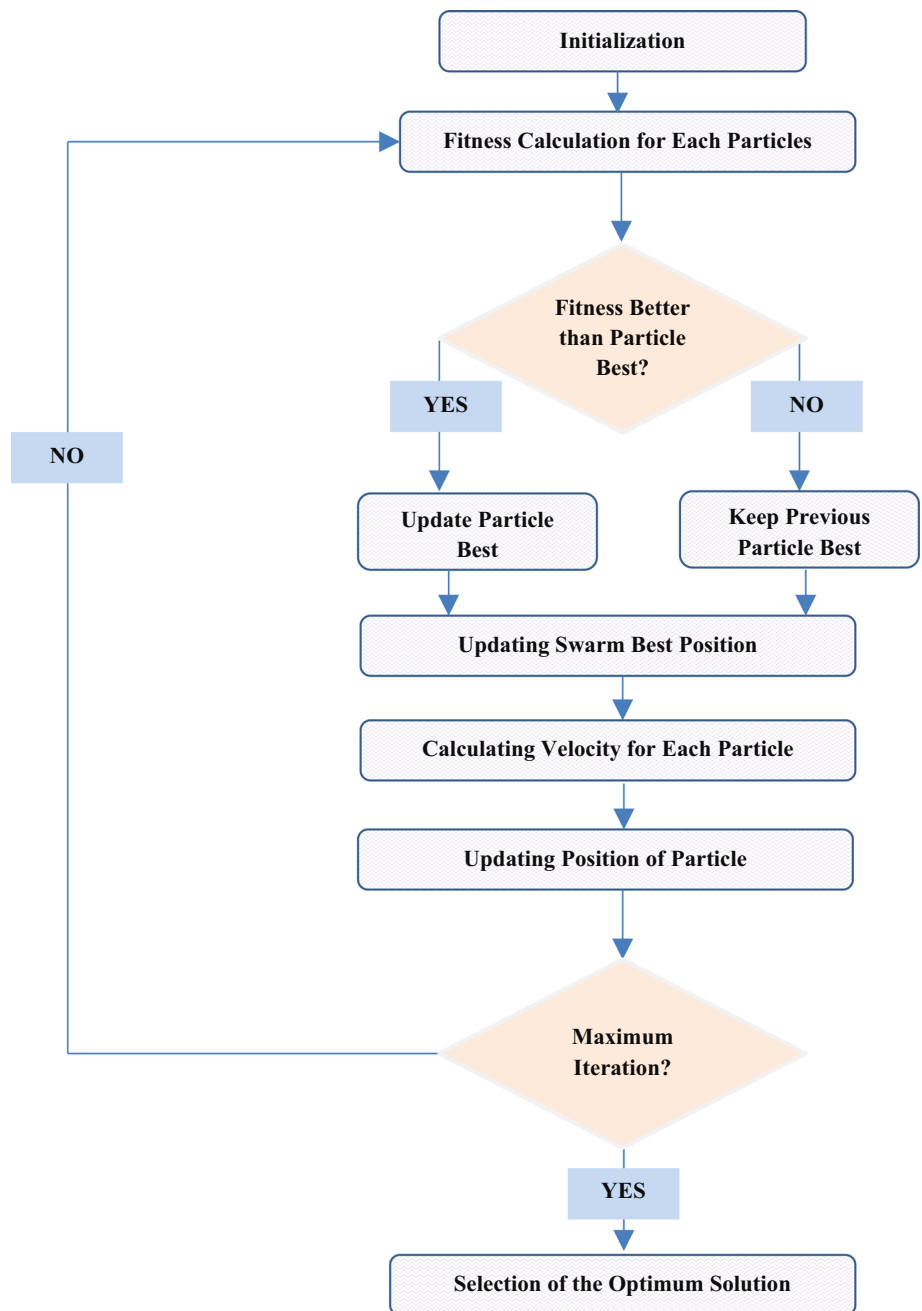
restaurant's services. In addition, customers' satisfaction is revealed through an overall rating in a range of 1 to 5. We collected 7562 ratings from TripAdvisor. The records with null values in four main criteria were removed. Accordingly, 3822 records remained in the dataset for further analysis.

In the first step, k-means clustering was applied on the data to obtain several clusters from the customers' ratings on the restaurants. In this study, totally, 6 clusters were generated from the data. Each cluster included the users' ratings on the Food, Service, Value and Atmosphere along with their overall ratings. In Fig. 5, we present the importance of factors in each cluster. Restaurants criteria in each cluster base on customers' ratings are presented in Table 4. In addition, customer satisfaction in each cluster is presented in Table 5. We present the results of k-means clustering in Fig. 6.

The clusters generated by k-means were used in PSO-ANN for predicting customers' satisfaction based on the inputs ratings. The data in each cluster were divided into training and test sets. The training sets (60% of the data) were used to construct the prediction models, and test sets (40% of
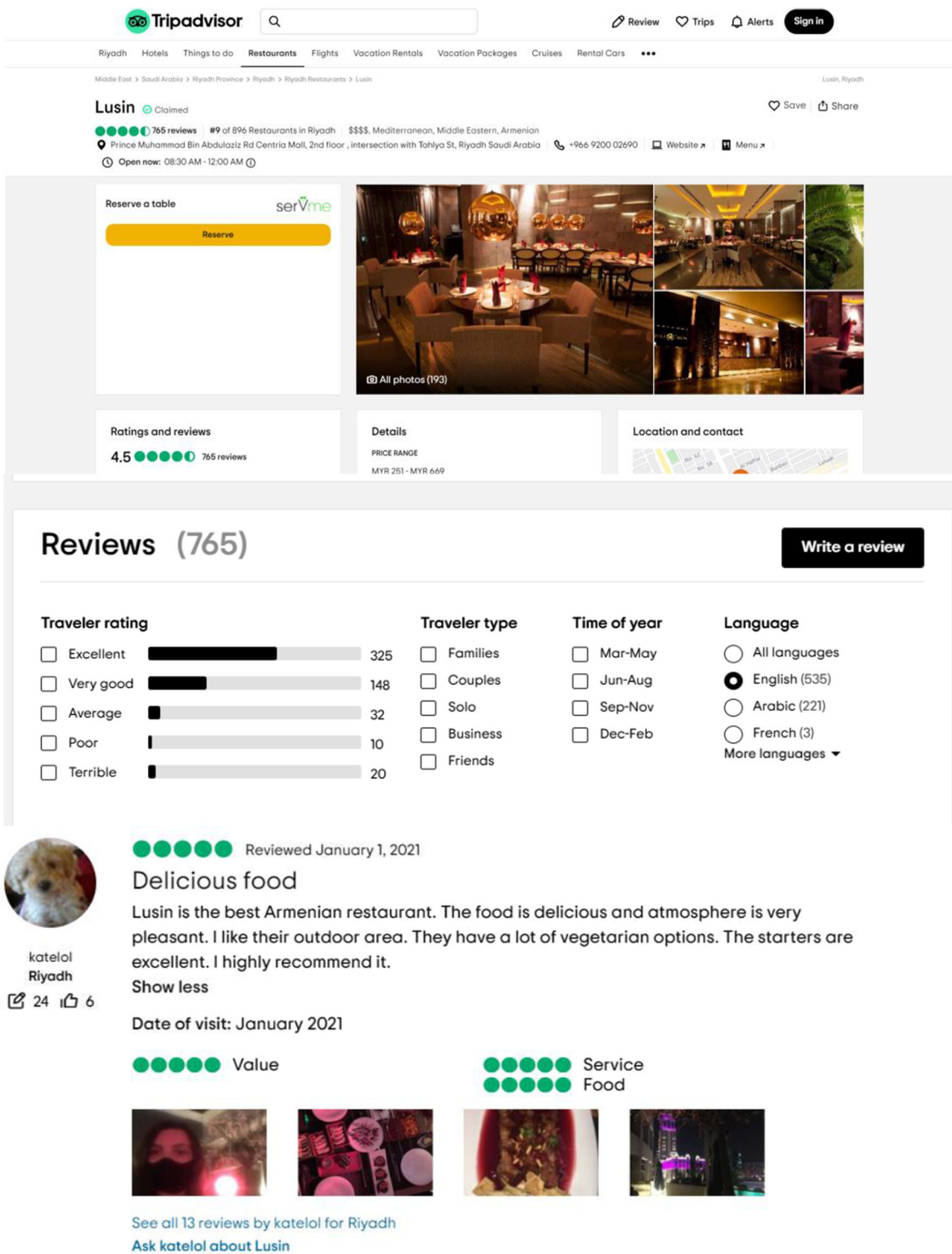
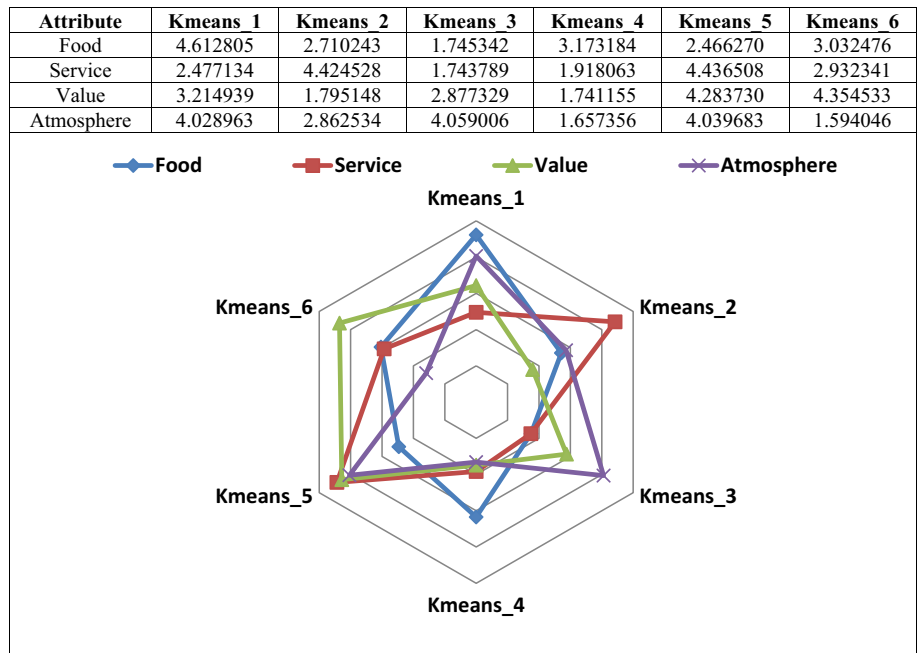**Fig. 4** Restaurants registered in TripAdvisor

Fig. 5 Cluster centroids

Attribute	Kmeans_1	Kmeans_2	Kmeans_3	Kmeans_4	Kmeans_5	Kmeans_6
Food	4.612805	2.710243	1.745342	3.173184	2.466270	3.032476
Service	2.477134	4.424528	1.743789	1.918063	4.436508	2.932341
Value	3.214939	1.795148	2.877329	1.741155	4.283730	4.354533
Atmosphere	4.028963	2.862534	4.059006	1.657356	4.039683	1.594046

Table 4 Restaurants criteria in each cluster

Criteria	Ratings	Clusters					
		Kmeans_1 Count	Kmeans_2 Count	Kmeans_3 Count	Kmeans_4 Count	Kmeans_5 Count	Kmeans_6 Count
Food	1	0	155	291	75	148	116
	2	0	186	226	96	134	160
	3	18	195	127	131	104	175
	4	218	131	0	131	75	160
	5	420	75	0	104	43	128
Service	1	160	0	279	198	0	120
	2	176	0	256	204	0	176
	3	184	89	104	116	63	179
	4	119	249	5	19	158	162
	5	17	404	0	0	283	102
Value	1	61	304	116	228	0	0
	2	142	286	154	220	0	0
	3	179	152	154	89	75	86
	4	143	0	133	0	211	305
	5	131	0	87	0	218	348
Atmosphere	1	0	111	0	246	0	365
	2	34	191	26	231	27	309
	3	157	222	146	58	120	65
	4	221	125	236	2	163	0
	5	244	93	236	0	194	0

Table 5 Customer satisfaction in each cluster

Criteria	Ratings	Clusters					
		Kmeans_1 Count	Kmeans_2 Count	Kmeans_3 Count	Kmeans_4 Count	Kmeans_5 Count	Kmeans_6 Count
Customer Satisfaction	1	0	37	70	144	0	24
	2	115	241	223	200	57	215
	3	195	255	237	169	147	287
	4	231	194	110	24	179	184
	5	115	15	4	0	121	29

the data) were used to evaluate the prediction models for their performance. To find the best network for this prediction, we performed an experiment on the use of different hidden layers of ANN and provided their mean squared error values. In this experiment, for different numbers of hidden layers in ANN, we present the swarm minimum mean squared error values versus iteration. According to this experiment, the swarm minimum Mean Squared Error values correspond to swarm best positions. The results are presented in Fig. 7. From this figure, it is found that a huge decrease in the swarm's minimum mean squared error occurred after one iteration. This indicates that PSO has been efficient in tuning the ANN architecture for customers' satisfaction prediction. In addition, as shown in Fig. 6, after the swarm' the minimum mean squared error was gradually reduced in the first 20 iterations. Overall, it was found that the lowest minimum error (MSE = 0.09190) was obtained by the ANN model in the first cluster with two hidden layers. In these two layers, eight neurons and two neurons, respectively, in the first and second layers were the best option. This experiment was performed for all 6 clusters to find an optimum ANN architecture, as shown in Fig. 7.

According to the optimum ANN models in six clusters, we present the results in Fig. 8 for the coefficient of determination. The results presented in the figures show that PSO-ANN has accurately predicted customers' satisfaction in six clusters. In addition, the results for the correlation coefficient in six segments show that a high coefficient of determination was achieved by PSO-ANN. In fact, the relationship between customers' satisfaction and quality dimensions has been successfully discovered by PSO-ANN. In Fig. 9, we present actual satisfaction versus predicted satisfaction in six clusters.

We also compared the results of PSO-ANN with ANN and SVR. The results of the comparisons are provided in Table 6. In this table, we have provided the average values of the MSE and coefficient of determination of the six models. The results clearly demonstrate that k-means-PSO-ANN (MSE = 0.09847; R^2 = 0.98764) has outperformed other methods, k-Means-ANN (MSE = 0.16134; R^2 = 0.86535), ANFIS

(MSE = 0.18829; R^2 = 0.87194), ANN Ensembles (MSE = 0.15632; R^2 = 0.94352), SVR (MSE = 0.18936; R^2 = 0.86535), k-Means-MLR (MSE = 0.18534; R^2 = 0.88342) and ANN (MSE = 0.17563; R^2 = 0.92678). This indicates that the use of PSO has improved the performance of ANN for customers' satisfaction analysis from TripAdvisor data. In addition, the results show that clustering has improved the prediction accuracy of the predictors.

5 Discussion

The competitiveness in different industries has become progressively important, especially for countries that rely mainly on tourism and hospitality. A hotel and restaurant can be assumed competitive if it can attract and satisfy pertinent consumers. The competitiveness of a destination not only has a direct effect on tourism revenue in terms of the number of visitors and expenses, but also indirectly influences tourism-related industries like hotels and restaurants [64]. User-generated data are considered as the main source of data for tourism and hospitality research and practices. They demonstrate customers' delight or dissatisfaction with their experiences [9]. In other words, user-generated data are easy and meaningful reaction created by customers that is openly attainable on the online setting [65]. User-generated data are the main approach that can benefit hospitality and tourism hence they help to improve competitiveness [66]. Online reviews have been considered as the primary channel for sharing feedback about different services. They support potential customers in the decision-making process and indirectly inspire hotel managers to upgrade the quality of their products or services [67]. The uniqueness of the hospitality and tourism products and services coupled with the growing importance of user-generated data has also triggered studies to determine the perspectives of hospitality customers, assisting to create managers' strategies [39]. Customer heterogeneity requires market segmentation to obtain suitable targeting, arranging, marketing and income management [68].

Fig. 6 Clustering of data by
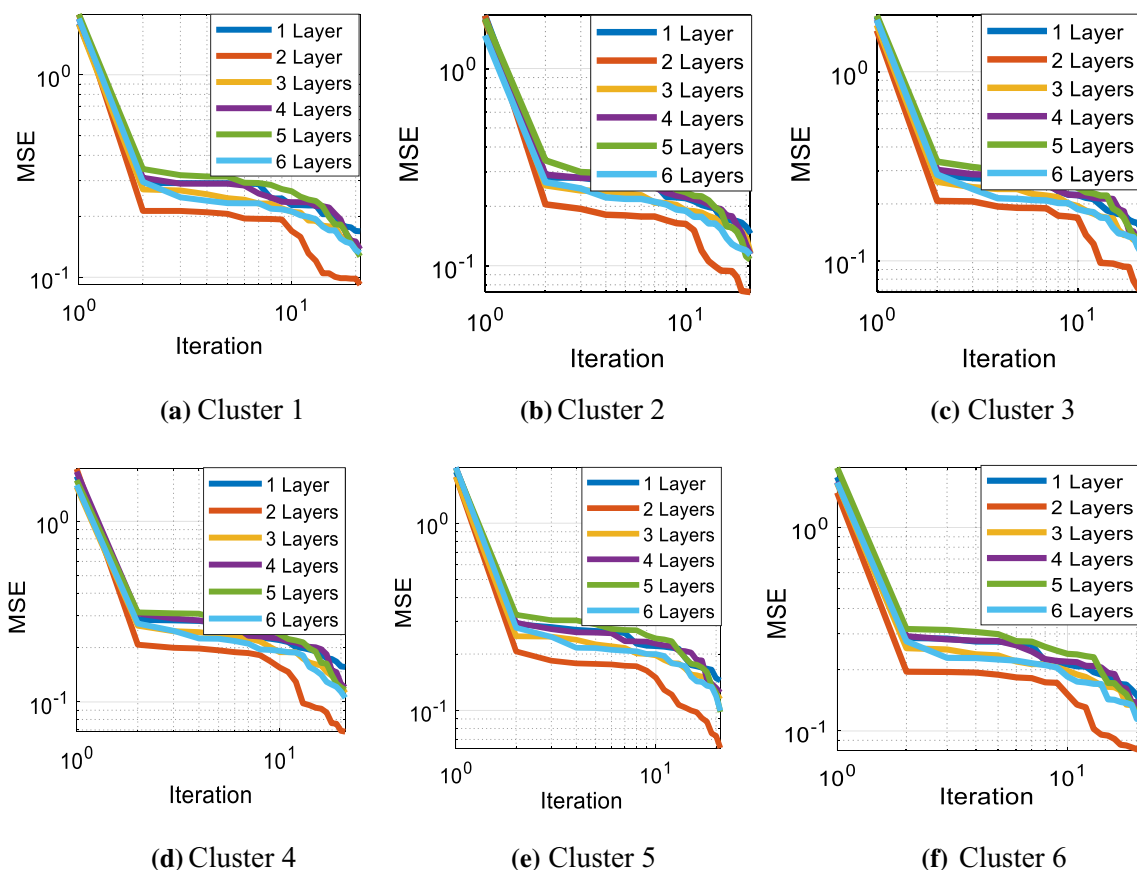*k*-means
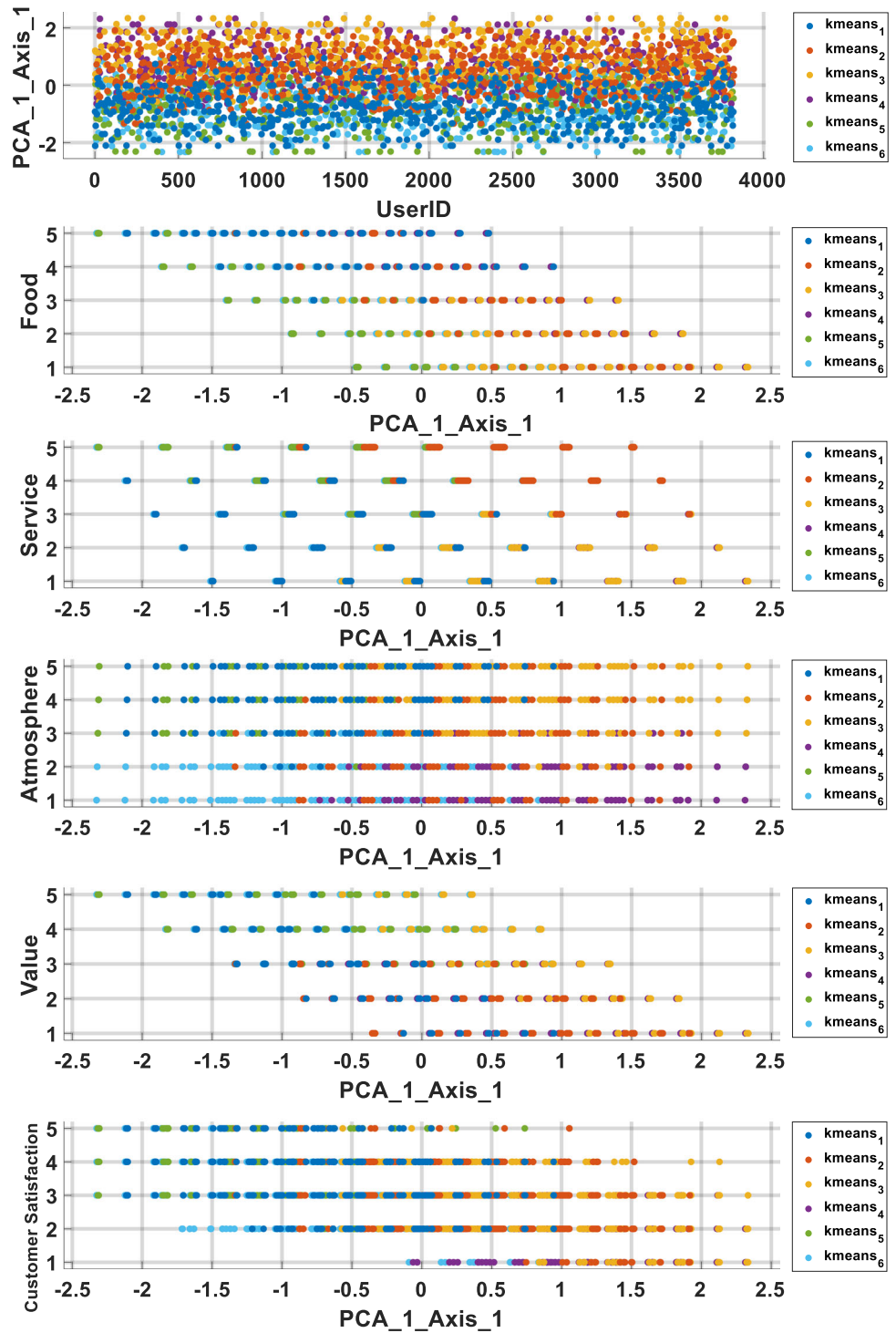
Fig. 7 MSE versus iteration in ANN for 6 clusters

When we have heterogeneous customers, allocating them into mutually exclusive and homogeneous clusters has proved to be a practical and so extensively used idea among marketing managers [3]. In the restaurant's context, customers usually find it difficult to select a restaurant according to elements like the taste, physical setting, price, and service quality [46]. Online users' efforts to create pertinent customers with reliable information related to the unfamiliar restaurant can diminish the problems and ambiguity of selecting a restaurant. Past researches have revealed the benefit of user-generated content for different services and products like hospitality services [44]. The literature demonstrated that service quality has a notable effect on user satisfaction and loyalty compared to product or service features. Quality of services also causes lowering costs, elevating profitability, increasing industry performance and finally positive word-of-mouth. Managers and policymakers have currently initiated to develop strategies to determine service quality to their superiority [69]. Furthermore, user satisfaction, which is considered as the overall evaluation of different features of

products or services, is one of the critical factors for business performance, competitiveness and profitability [70].

Past researchers have revealed that in the hospitality industry, the accurate measurement of a company's success lies in its capability to satisfy customer's preferences consistently. In other words, it is ascertained that service quality is antecedent to customer satisfaction [71]. In this way, many researchers are aiming to determine customer insight regarding products or services provided in hotels and restaurants and the impact of this insight on customer satisfaction [30, 44, 70–72]. On the other hand, social media has dramatically changed the way industries communicate with customers. The main characteristic of today's online business is that customers use social media to develop and maintain communication with a vast group of individuals [15]. Studies explain that eWOM communication on social media is one of the most impressive channels to set up consumer attitudes regarding a product or service and these contents have a critical impact on customer purchase decisions [4, 19, 35, 39, 48, 49, 73].

This study contributed to the analysis of social data provided by restaurants' customers using a hybrid method. The customer segmentation in Saudi Arabia restaurants was performed using clustering, neural networks, and optimization learning techniques. We collected 3822 ratings from TripAdvisor for data analysis to find the relationship between the quality dimension of restaurants and customers' satisfaction. In the first step, k-means clustering was applied on the data to obtain several clusters from the customers' ratings on the restaurants. Totally, six clusters were generated from the TripAdvisor data. Each cluster included the users' ratings on the Food, Service, Value and Atmosphere along with their overall ratings. The importance of the factors impacting customers' satisfaction was revealed in each cluster. It was found that in Cluster 1, Food, Atmosphere and Value were the most important factors. In Cluster 2, Service was highly impacted on the customers' satisfaction. In Cluster 3, Atmosphere was important for the customers and Food and Services were their main concerns. In Cluster 4, Food has highly influenced the customers' satisfaction. Service, Value and Atmosphere were the most important factors in Cluster 5. Food and Value were highly impacted on the customers' satisfaction in Cluster 6. These clusters provide valuable information about customers' preferences in Saudi Arabia restaurants. In addition, in this study, we have collected real data for our method evaluation. Nowadays, the analysis of social data using sophisticated learning approaches can aid business managers in better detecting the customers' segments. This will accordingly improve their competitive advantages in the market.

6 Conclusions

This study developed a new method for analyzing customers' satisfaction in Saudi Arabia restaurants through the use of social data. We used clustering and ANN to develop the hybrid method. The performance of the ANN technique was improved using PSO. We collected data from TripAdvisor to investigate the customers' satisfaction in the restaurants. The effectiveness of the method was measured using MSE and coefficient of determination values. In addition, we compared the results with other methods. The results demonstrated that PSO-ANN (MSE $= 0.09847$; $R^2 = 0.98764$) has outperformed other methods, k-means-PSO-ANN (MSE $= 0.09847$; $R^2 = 0.98764$) has outperformed other methods, k-Means-ANN (MSE $= 0.16134$; $R^2 = 0.86535$), SVR (MSE $= 0.18936$; $R^2 = 0.86535$), k-Means-MLR (MSE $= 0.18534$; $R^2 = 0.88342$) and ANN (MSE $= 0.17563$; $R^2 = 0.92678$), in analyzing customers' satisfaction. In future research, it
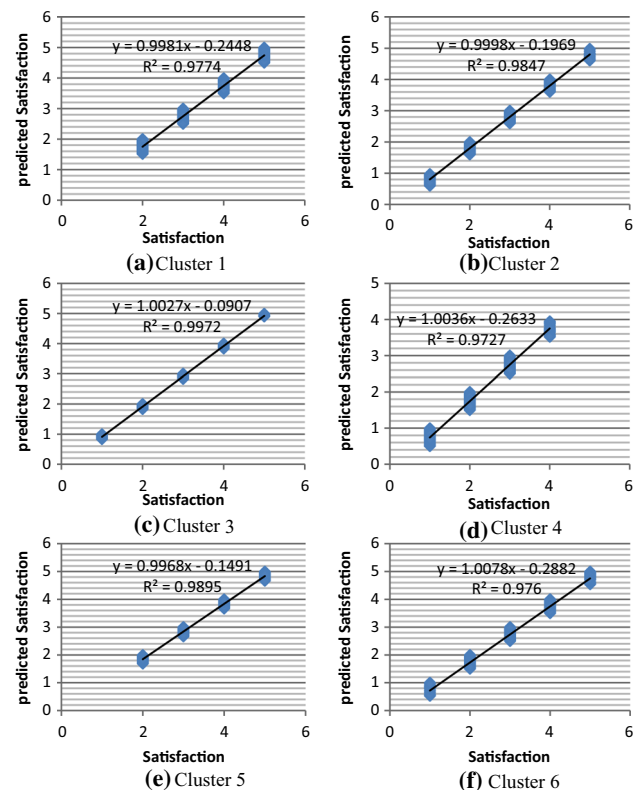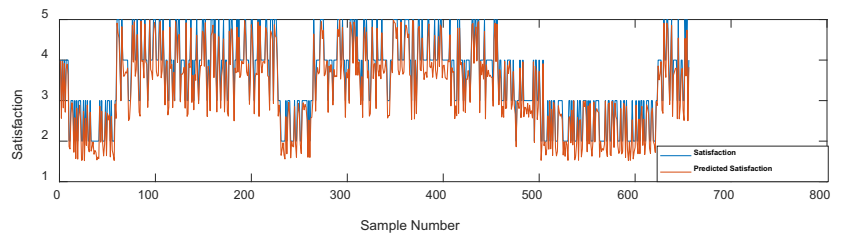
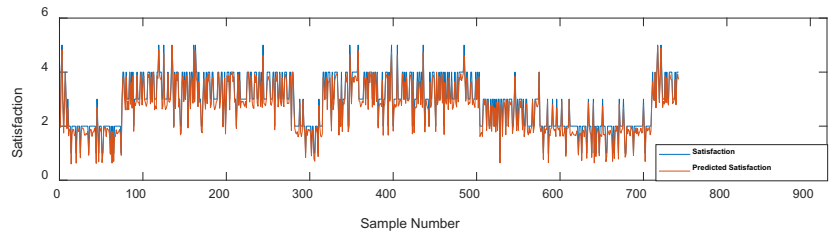**Fig. 8** Coefficient of determination in six clusters

is recommended that different sources of data are used for method evaluation.

A large-scale study is a complementary approach to obtain outcomes that can significantly advance existing knowledge in customers' preference learning and segmentation and can be used in conjunction with other approaches. Accordingly, compared to a small study with a limited number of participants or a small sample size, a large study with a very large population size is more likely to be successful in this context. In fact, large samples can produce estimates of effects with high precision and provide the capability to better detect the segments. In addition, as SVR has also provided accurate results, its effectiveness could be investigated using optimization techniques for analyzing customers' satisfaction. Furthermore, this research did not include textual reviews for customers' satisfaction assessments. Therefore, the text mining approaches can be incorporated into the proposed method for customers satisfaction analysis and segmentation. Moreover, the proposed method can be further developed by new machine learning algorithms [58, 74–78] and compared with the proposed method to see its effectiveness in term of prediction accuracy. Additionally, robust imputation methods for missing data will be effective in improving the accuracy of the prediction learning techniques [79, 80]. Finally, we used the NN technique for the evaluation of customers' satisfaction using social data, however, Explainable NN (XNN)
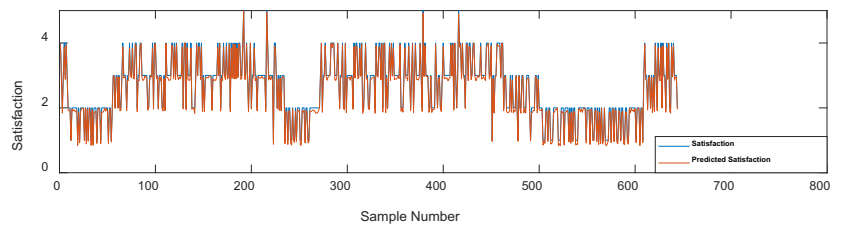
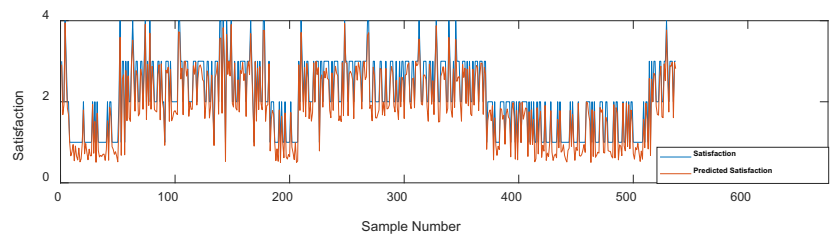**Fig. 9** Actual satisfaction versus predicted satisfaction in six clusters
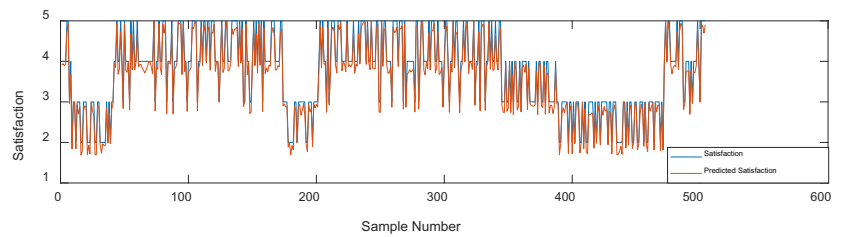

(a) Cluster 1

(b) Cluster 2

(c) Cluster 3

(d) Cluster 4

(e) Cluster 5
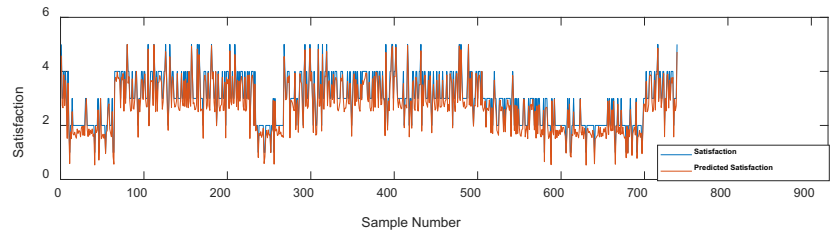
(f) Cluster 6

Table 6 Method comparisons

Method	MSE	Coefficient of Determination (R^2)
k-Means-PSO-ANN	0.09847	0.98764
ANN	0.18936	0.86535
SVR	0.17563	0.91234
k-Means-ANN	0.16134	0.92678
k-Means-MLR	0.18534	0.88342
ANFIS	0.18829	0.87194
ANN Ensembles	0.15632	0.94352

could be used for the explanation of the features extracted by different layers of the network [81–83].

References

1. DiPietro, R.B.; Levitt, J.A.; Taylor, S.; Nierop, T.: First-time and repeat tourists' perceptions of authentic Aruban restaurants: an importance-performance competitor analysis. J. Destin. Mark. Manag. **14**, 100366 (2019)
2. Line, N.D.; Hanks, L.; Dogru, T.: A reconsideration of the EWOM construct in restaurant research: what are we really measuring? Int. J. Contemp. Hosp. Manag. (2020). https://doi.org/10.1108/IJCHM-06-2020-0561
3. Arunachalam, D.; Kumar, N.: Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making. Expert Syst. Appl. **111**, 11–34 (2018)
4. Leung, D.; Law, R.; Van Hoof, H.; Buhalis, D.: Social media in tourism and hospitality: a literature review. J. Travel Tour. Mark. **30**(1–2), 3–22 (2013)
5. Amatulli, C.; De Angelis, M.; Stoppani, A.: Analyzing online reviews in hospitality: data-driven opportunities for predicting the sharing of negative emotional content. Curr. Issue Tour. **22**(15), 1904–1917 (2019)
6. Zheng, T.; Wu, F.; Law, R.; Qiu, Q.; Wu, R.: Identifying unreliable online hospitality reviews with biased user-given ratings: a deep learning forecasting approach. Int. J. Hosp. Manag. **92**, 102658 (2021)
7. Aleti, T.; Pallant, J.I.; Tuan, A.; van Laer, T.: Tweeting with the stars: automated text analysis of the effect of celebrity social media communications on consumer word of mouth. J. Interact. Mark. **48**, 17–32 (2019)
8. Nakayama, M.; Wan, Y.: The cultural impact on social commerce: a sentiment analysis on Yelp ethnic restaurant reviews. Inform. Manag. **56**(2), 271–279 (2019)
9. Taecharungroj, V.; Mathayomchan, B.: Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tour. Manag. **75**, 550–568 (2019)
10. Nilashi, M.; Minaei-Bidgoli, B.; Alrizq, M.; Alghamdi, A.; Alsulami, A.A.; Samad, S.; Mohd, S.: An analytical approach for big social data analysis for customer decision-making in eco-friendly hotels. Expert Syst. Appl. **186**, 115722 (2021)
11. Talón-Ballestero, P.; González-Serrano, L.; Soguero-Ruiz, C.; Munoz-Romero, S.; Rojo-Álvarez, J.L.: Using big data from customer relationship management information systems to determine the client profile in the hotel sector. Tour. Manag. **68**, 187–197 (2018)
12. Barkah, C.S.; Raharja, S.J.: Analysis of marketing strategies and competitive advantages of two producers special foods in garut. Int. J. Trade Global Mark. **11**(1–2), 87–97 (2018)
13. Dursun, A.; Caber, M.: Using data mining techniques for profiling profitable hotel customers: an application of RFM analysis. Tour. Manag. Perspect. **18**, 153–160 (2016)
14. Ahani, A.; Nilashi, M.; Yadegaridehkordi, E.; Sanzogni, L.; Tarik, A.R.; Knox, K.; Samad, S.; Ibrahim, O.: Revealing customers' satisfaction and preferences through online review analysis: the case of Canary Islands hotels. J. Retail. Consum. Serv. **51**, 331–343 (2019)
15. Kim, W.G.; Lim, H.; Brymer, R.A.: The effectiveness of managing social media on hotel performance. Int. J. Hosp. Manag. **44**, 165–171 (2015)
16. Zhang, C.; Zhang, H.; Wang, J.: Personalized restaurant recommendation method combining group correlations and customer preferences. Inf. Sci. **454–455**, 128–143 (2018)
17. Nilashi, M.; Ahmadi, H.; Arji, G.; Alsalem, K.O.; Samad, S.; Ghabban, F.; Alzahrani, A.O.; Ahani, A.; Alarood, A.A.: Big social data and customer decision making in vegetarian restaurants: a combined machine learning method. J. Retail. Consum. Serv. **62**, 102630 (2021)
18. An, J.; Kwak, H.; Jung, S.-G.; Salminen, J.; Jansen, B.J.: Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Soc. Netw. Anal. Min. **8**(1), 1–19 (2018)
19. Müller, J.M.; Pommeranz, B.; Weisser, J.; Voigt, K.-I.: Digital, social media, and mobile marketing in industrial buying: still in need of customer segmentation? Empirical evidence from Poland and Germany. Ind. Mark. Manag. **73**, 70–83 (2018)
20. Li, J.; Yang, Y.; Baibokonov, D.: Research on Customer Classification and Service Quality Evaluation of Online Education Platform. (2020)
21. Aksu, A.; Albayrak, T.; Caber, M.: Hotel customer segmentation according to eco-service quality perception: the case of Russian tourists. J. Hosp. Tour. Insights (2021). https://doi.org/10.1108/JHTI-10-2020-0188
22. Legohérel, P.; Hsu, C.H.; Daucé, B.: Variety-seeking: Using the CHAID segmentation approach in analyzing the international traveler market. Tour. Manag. **46**, 359–366 (2015)
23. Nilashi, M.; Ahani, A.; Esfahani, M.D.; Yadegaridehkordi, E.; Samad, S.; Ibrahim, O.; Sharef, N.M.; Akbari, E.: Preference learning for eco-friendly hotels recommendation: a multi-criteria collaborative filtering approach. J. Clean. Prod. **215**, 767–783 (2019)
24. Lee, F.S.J.: Hospitality products and the consumer price-perceived quality heuristic: an empirical perspective. Serv. Mark. Q. **34**(3), 205–214 (2013)
25. Lu, L.; Gursoy, D.: Does offering an organic food menu help restaurants excel in competition? An examination of diners' decision-making. Int. J. Hosp. Manag. **63**, 72–81 (2017)
26. Melina, V.; Craig, W.; Levin, S.: Position of the academy of nutrition and dietetics: vegetarian diets. J. Acad. Nutr. Diet. **116**(12), 1970–1980 (2016)
27. Cheng, C.-H.; Lin, S.-Y.; Tsai, C.-C.: Investigating consumer preferences in choosing vegetarian restaurants using conjoint analysis. Curr. Urban Stud. **2**(03), 279 (2014)
28. Trappey, C.V.; Trappey, A.J.C.; Chang, A.C.; Huang, A.Y.L.: The analysis of customer service choices and promotion preferences using hierarchical clustering. J. Chin. Inst. Ind. Eng. **26**(5), 367–376 (2009)

29. Nilashi, M.; Mardani, A.; Liao, H.; Ahmadi, H.; Manaf, A.A.; Almukadi, W.: A hybrid method with TOPSIS and machine learning techniques for sustainable development of green hotels considering online reviews. Sustainability **11**(21), 6013 (2019)

30. Sánchez-Franco, M.J.; Navarro-García, A.; Rondán-Cataluña, F.J.: A naive Bayes strategy for classifying customer satisfaction: a study based on online reviews of hospitality services. J. Bus. Res. **101**, 499–506 (2019)

31. Yadegaridehkordi, E.; Nilashi, M.; Nasir, M.H.N.B.M.; Momtazi, S.; Samad, S.; Supriyanto, E.; Ghabban, F.: Customers segmentation in eco-friendly hotels using multi-criteria and machine learning techniques. Technol. Soc. **65**, 101528 (2021)

32. Kazbare, L.; van Trijp, H.C.; Eskildsen, J.K.: A-priori and post-hoc segmentation in the design of healthy eating campaigns. J. Mark. Commun. **16**(1–2), 21–45 (2010)

33. Gao, S.; Tang, O.; Wang, H.; Yin, P.: Identifying competitors through comparative relation mining of online reviews in the restaurant industry. Int. J. Hosp. Manag. **71**, 19–32 (2018)

34. Kwok, L.; Mao, Z.; Huang, Y.K.: Consumers' electronic word-of-mouth behavioral intentions on Facebook: does message type have an effect? Tour. Hosp. Res. **19**(3), 296–307 (2019)

35. Xiang, Z.; Du, Q.; Ma, Y.; Fan, W.: A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour. Manag. **58**, 51–65 (2017)

36. Zhang, Y.; Feng, X.; Wu, Y.; Xiao, J.: The dynamics of online ratings with heterogeneous preferences in online review platform. Chaos Solitons Fractals **109**, 26–30 (2018)

37. Ariyasriwatana, W.; Buente, W.; Oshiro, M.; Streveler, D.: Categorizing health-related cues to action: using Yelp reviews of restaurants in Hawaii. New Rev. Hypermedia Multimedia **20**(4), 317–340 (2014)

38. Khorsand, R.; Rafiee, M.; Kayvanfar, V.: Insights into TripAdvisor's online reviews: the case of Tehran's hotels. Tour. Manag. Perspect. **34**, 100673 (2020)

39. Kim, W.G.; Li, J.; Brymer, R.A.: The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. Int. J. Hosp. Manag. **55**, 41–51 (2016)

40. Moro, S.; Rita, P.; Coelho, J.: Stripping customers' feedback on hotels through data mining: the case of Las Vegas Strip. Tour. Manag. Perspect. **23**, 41–52 (2017)

41. Park, E.O.; Chae, B.K.; Kwon, J.: The structural topic model for online review analysis: comparison between green and non-green restaurants. J. Hosp. Tour. Technol. (2018). https://doi.org/10.1108/JHTT-08-2017-0075

42. Basri, N.A.M.H.; Ahmad, R.; Anuar, F.I.; Ismail, K.A.: Effect of word of mouth communication on consumer purchase decision: malay upscale restaurant. Procedia Soc. Behav. Sci. **222**, 324–331 (2016)

43. Gan, Q.; Ferns, B.H.; Yu, Y.; Jin, L.: A text mining and multidimensional sentiment analysis of online restaurant reviews. J. Qual. Assur. Hosp. Tour. **18**(4), 465–492 (2017)

44. Konuk, F.A.: The influence of perceived food quality, price fairness, perceived value and satisfaction on customers' revisit and word-of-mouth intentions towards organic food restaurants. J. Retail. Consum. Serv. **50**, 103–110 (2019)

45. Nilashi, M.; Samad, S.; Ahani, A.; Ahmadi, H.; Alsolami, E.; Mahmoud, M.; Majeed, H.D.; Alarood, A.A.: Travellers decision making through preferences learning: a case on Malaysian spa hotels in TripAdvisor. Comput. Ind. Eng. **158**, 107348 (2021)

46. Zhang, Z.; Ye, Q.; Law, R.; Li, Y.: The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int. J. Hosp. Manag. **29**(4), 694–700 (2010)

47. Khalili-Damghani, K.; Abdi, F.; Abolmakarem, S.: Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case

48. of customer-centric industries. Appl. Soft Comput. **73**, 816–828 (2018)

48. Aydin, G.: Social media engagement and organic post effectiveness: a roadmap for increasing the effectiveness of social media use in hospitality industry. J. Hosp. Market. Manag. **29**(1), 1–21 (2020)

49. Mavragani, E.; Nikolaidou, P.; Theodoraki, E.: Traveler segmentation through Social Media for intercultural marketing purposes: the case of Halkidiki. J. Tour. Herit. Serv. Mark. **5**(1), 15–23 (2019)

50. Dolnicar, S.: Market Segmentation in Tourism: Managing Tourism Firms. Edward Elgar Publishing, Cheltenham (2008)

51. Fernandes, E.; Moro, S.; Cortez, P.; Batista, F.; Ribeiro, R.: A data-driven approach to measure restaurant performance by combining online reviews with historical sales data. Int. J. Hosp. Manag. **94**, 102830 (2021)

52. Prasad, N.N.; Rao, J.N.: The estimation of the mean squared error of small-area estimators. J. Am. Stat. Assoc. **85**(409), 163–171 (1990)

53. Srivastava, A.K.; Srivastava, V.K.; Ullah, A.: The coefficient of determination and its adjusted version in linear regression models. Econom. Rev. **14**(2), 229–240 (1995)

54. Anitha, P.; Patil, M.M.P: RFM model for customer purchase behavior using K-Means algorithm. J. King Saud Univ. Comput. Inform. Sci. (2019)

55. Chiu, C.-Y.; Chen, Y.-F.; Kuo, I.-T.; Ku, H.C.: An intelligent market segmentation system using k-means and particle swarm optimization. Expert Syst. Appl. **36**(3), 4558–4565 (2009)

56. Ye, L.; Qiu-ru, C.; Hai-xu, X.; Yi-jun, L.; Zhi-min, Y.: Telecom customer segmentation with K-means clustering. In: 2012 7th International Conference on Computer Science & Education (ICCSE). IEEE, pp. 648–651. (2012)

57. Zare, H.; Emadi, S.: Determination of customer satisfaction using Improved K-means algorithm. Soft. Comput. **24**, 16947–16965 (2020)

58. Zhang, W.; Li, H.; Li, Y.; Liu, H.; Chen, Y.; Ding, X.: Application of deep learning algorithms in geotechnical engineering: a short critical review. Artif. Intell. Rev. **54**(8), 5633–5673 (2021)

59. Zhang, W.; Zhang, R.; Wu, C.; Goh, A.T.C.; Lacasse, S.; Liu, Z.; Liu, H.: State-of-the-art review of soft computing applications in underground excavations. Geosci. Front. **11**(4), 1095–1106 (2020)

60. Zhang, W.; Wu, C.; Li, Y.; Wang, L.; Samui, P.: Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. Georisk Assess. Manag. Risk Eng. Syst. Geohazards **15**(1), 27–40 (2021)

61. Chandrasekaran, M.; Tamang, S.: ANN–PSO integrated optimization methodology for intelligent control of MMC machining. J. Inst. Eng. India Ser. **98**(4), 395–401 (2017)

62. Da, Y.; Xiurun, G.: An improved PSO-based ANN with simulated annealing technique. Neurocomputing **63**, 527–533 (2005)

63. Gambhir, S.; Malik, S.K.; Kumar, Y.: PSO-ANN based diagnostic model for the early detection of dengue disease. New Horiz. Transl. Med. **4**(1–4), 1–8 (2017)

64. Tsai, H.; Song, H.; Wong, K.K.: Tourism and hotel competitiveness research. J. Travel Tour. Mark. **26**(5–6), 522–546 (2009)

65. Guo, Y.; Barnes, S.J.; Jia, Q.: Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tour. Manag. **59**, 467–483 (2017)

66. Simeon, M.I.; Buonincontri, P.; Cinquegrani, F.; Martone, A.: Exploring tourists' cultural experiences in Naples through online reviews. J. Hosp. Tour. Technol. (2017). https://doi.org/10.1108/JHTT-10-2016-0067

67. Martinez-Torres, M.R.; Toral, S.L.: A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tour. Manag. **75**, 393–403 (2019)

68. Nessel, K.; Kościółek, S.; Wszendybył-Skulska, E.; Kopera, S.: Benefit segmentation in the tourist accommodation market based on eWOM attribute ratings. Inform. Technol. Tour. **23**(2), 265–290 (2021)

69. Tontini, G.; dos Santos Bento, G.; Milbratz, T.C.; Volles, B.K.; Ferrari, D.: Exploring the nonlinear impact of critical incidents on customers' general evaluation of hospitality services. Int. J. Hosp. Manag. **66**, 106–116 (2017)

70. Davras, Ö.; Caber, M.: Analysis of hotel services by their symmetric and asymmetric effects on overall customer satisfaction: a comparison of market segments. Int. J. Hosp. Manag. **81**, 83–93 (2019)

71. Amin, M.; Yahya, Z.; Ismayatim, W.F.A.; Nasharuddin, S.Z.; Kassim, E.: Service quality dimension and customer satisfaction: an empirical study in the Malaysian hotel industry. Serv. Mark. Q. **34**(2), 115–125 (2013)

72. Lee, Y.K.; Park, K.H.; Park, D.H.; Lee, K.A.; Kwon, Y.J.: The relative impact of service quality on service value, customer satisfaction, and customer loyalty in Korean family restaurant context. Int. J. Hosp. Tour. Adm. **6**(1), 27–51 (2005)

73. Standing, C.; Holzweber, M.; Mattsson, J.: Exploring emotional expressions in e-word-of-mouth from online communities. Inf. Process. Manag. **52**(5), 721–732 (2016)

74. Chen, J.; Yang, G.; Khan, H.; Zhang, H.; Zhang, Y.; Zhao, S.; Mohiaddin, R.; Wong, T.; Firmin, D.; Keegan, J.: JAS-GAN: generative adversarial network based joint atrium and scar segmentations on unbalanced atrial targets. IEEE J. Biomed. Health Inform. **26**(1), 103–114 (2021)

75. Jin, Y.; Yang, G.; Fang, Y.; Li, R.; Xu, X.; Liu, Y.; Lai, X.: 3D PBV-Net: an automated prostate MRI data segmentation method. Comput. Biol. Med. **128**, 104160 (2021)

76. Liu, Y.; Yang, G.; Hosseiny, M.; Azadikhah, A.; Mirak, S.A.; Miao, Q.; Raman, S.S.; Sung, K.: Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. IEEE Access **8**, 151817–151828 (2020)

77. Yang, G.; Chen, J.; Gao, Z.; Li, S.; Ni, H.; Angelini, E.; Wong, T.; Mohiaddin, R.; Nyktari, E.; Wage, R.: Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. Futur. Gener. Comput. Syst. **107**, 215–228 (2020)

78. Zhang, W.; Wu, C.; Zhong, H.; Li, Y.; Wang, L.: Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. **12**(1), 469–477 (2021)

79. Choudhury, S.J.; Pal, N.R.: Imputation of missing data with neural networks for classification. Knowl.Based Syst. **182**, 104838 (2019)

80. Liu, T.; Wei, H.; Zhang, K.: Wind power prediction with missing data using Gaussian process regression and multiple imputation. Appl. Soft Comput. **71**, 905–916 (2018)

81. Vaughan, J.; Sudjianto, A.; Brahimi, E.; Chen, J.; Nair, V.N.: Explainable Neural Networks Based on Additive Index Models. (2018). arXiv preprint arXiv:1806.01933.

82. Yang, G.; Ye, Q.; Xia, J.: Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. Inf. Fusion **77**, 29–52 (2022)

83. Ye, Q.; Xia, J.; Yang, G.: Explainable AI for COVID-19 CT classifiers: an initial comparison study. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). IEEE, pp. 521–526. (2021)

84. Wang, Y.; Kim, J.: Interconnectedness between online review valence, brand, and restaurant performance. J. Hosp. Tour. Manag. **48**, 138–145 (2021)

85. Tian, G.; Lu, L.; McIntosh, C.: What factors affect consumers' dining sentiments and their ratings: evidence from restaurant online review data. Food Qual. Prefer. **88**, 104060 (2021)

86. Li, H.; Xie, K.L.; Zhang, Z.: The effects of consumer experience and disconfirmation on the timing of online review: field evidence from the restaurant business. Int. J. Hosp. Manag. **84**, 102344 (2020)

87. Tan, H.; Lv, X.; Liu, X.; Gursoy, D.: Evaluation nudge: effect of evaluation mode of online customer reviews on consumers' preferences. Tour. Manag. **65**, 29–40 (2018)

88. Berezan, O.; Raab, C.; Tanford, S.; Kim, Y.: Evaluating loyalty constructs among hotel reward program members using eWom. J. Hosp. Tour. Res. **39**(2), 198–224 (2015)

89. Weber, K.; Bradley, G.L.; Sparks, B.: Stressor effects of negative online reviews on anger and burnout in the restaurant industry. Int. J. Contemp. Hosp. Manag. **29**(11), 2847–2866 (2017)

90. Xie, K.L.; Chen, C.; Wu, S.: Online consumer review factors affecting offline hotel popularity: evidence from tripadvisor. J. Travel Tour. Mark. **33**(2), 211–223 (2016)

91. Kim, E.E.K.: The impact of restaurant service experience valence and purchase involvement on consumer motivation and intention to engage in eWOM. J. Qual. Assur. Hosp. Tour. **18**(3), 259–281 (2017)